

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:00:23 ; Search time 2331.3 Seconds  
(without alignments)  
1578.034 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416  
Perfect score: 223  
Sequence: 1 ctaccataattataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*\*  
1: gb\_ba:\*\*  
2: gb\_hgt:\*\*  
3: gb\_in:\*\*  
4: gb\_om:\*\*  
5: gb\_ov:\*\*  
6: gb\_pat:\*\*  
7: gb\_ph:\*\*  
8: gb\_pl:\*\*  
9: gb\_pr:\*\*  
10: gb\_ro:\*\*  
11: gb\_sts:\*\*  
12: gb\_sy:\*\*  
13: gb\_un:\*\*  
14: gb\_vi:\*\*  
15: em\_ba:\*\*  
16: em\_fun:\*\*  
17: em\_hum:\*\*  
18: em\_in:\*\*  
19: em\_om:\*\*  
20: em\_or:\*\*  
21: em\_ov:\*\*  
22: em\_pat:\*\*  
23: em\_ph:\*\*  
24: em\_pl:\*\*  
25: em\_ro:\*\*  
26: em\_sts:\*\*  
27: em\_sy:\*\*  
28: em\_un:\*\*  
29: em\_vi:\*\*  
30: em\_htgo\_hum:\*\*  
31: em\_htgo\_inv:\*\*  
32: em\_htgo\_rod:\*\*  
33: em\_htg\_hum:\*\*  
34: em\_htg\_inv:\*\*  
35: em\_htg\_rod:\*\*  
36: em\_htg\_other:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	223	100.0	416	6	AX127757	AX127757 Sequence
2	189.8	85.1	415	6	AX127755	AX127755 Sequence
3	47.8	21.4	186185	9	AC008817	AC008817 Homo sapi
4	46.4	20.8	110434	9	HSU66722	U66722 Human chrom
5	46	20.6	138350	2	AC010801	AC010801 Homo sapi
6	46	20.6	142126	2	AL596266	AL596266 Homo sapi
7	45.6	20.4	81503	2	AC091684	AC091684 Homo sapi
8	45.2	20.3	146349	2	AC007863	AC007863 Trypanoso
9	45.2	20.3	149241	2	AC068887	AC068887 Homo sapi
10	44.2	19.8	158236	9	AC011631	AC011631 Homo sapi
11	44.2	19.8	216431	2	AL354771	AL354771 Homo sapi
12	44	19.7	40917	6	AX033911	AX033911 Sequence
13	44	19.7	41008	6	AX033912	AX033912 Sequence
14	44	19.7	80155	9	AF042484	AF042484 Homo sapi
15	43.8	19.7	80622	9	AF222855	AF222855 Homo sapi
16	43.8	19.6	175959	2	AL450429	AL450429 Homo sapi
17	43.6	19.6	228	8	AY020821	AY020821 Oryza sat
18	43.6	19.6	234	8	AY018525	AY018525 Oryza sat
19	43.6	19.6	185499	2	AC090578	AC090578 Homo sapi
20	43.6	19.6	187967	2	AC026089	AC026089 Homo sapi
21	43.4	19.5	145358	2	AC091086	AC091086 Oryza sat
22	43.2	19.4	108523	9	AC000159	AC000159 Homo sapi
23	43.2	19.4	135237	2	AP000814	AP000814 Homo sapi
24	43.2	19.4	156736	9	AC006502	AC006502 Homo sapi
25	43	19.3	191962	2	AC048342	AC048342 Homo sapi
26	43	19.3	193996	2	AC068503	AC068503 Homo sapi
27	43	19.3	203718	9	AC009486	AC009486 Homo sapi
28	42.8	19.2	156060	2	AC004153	AC004153 Plasmodi
29	42.8	19.2	165114	2	AC011854	AC011854 Homo sapi
30	42.8	19.2	175020	9	AC022317	AC022317 Homo sapi
31	42.8	19.2	185702	2	AP002963	AP002963 Homo sapi
32	42.8	19.2	211316	9	AC009294	AC009294 Homo sapi
33	42.8	19.2	269661	2	AC023911	AC023911 Homo sapi
34	42.4	19.0	69058	9	AF222856	AF222856 Homo sapi
35	42.4	19.0	80202	9	AF222854	AF222854 Homo sapi
36	42.4	19.0	101584	9	CNS01DS5	AL121655 BAC seque
37	42.4	19.0	155390	2	AL596118	AL596118 Homo sapi
38	42.4	19.0	162958	2	AC011022	AC011022 Homo sapi
39	42.4	19.0	168519	2	AL355340	AL355340 Homo sapi
40	42.4	19.0	168671	2	AC019247	AC019247 Homo sapi
41	42.4	19.0	173435	2	AC015715	AC015715 Homo sapi
42	42.4	19.0	176432	2	AL391648	AL391648 Homo sapi
43	42.4	19.0	180942	33	AL512845	AL512845 Human DNA
44	42.2	18.9	66623	2	AC073039	AC073039 Homo sapi
45	42.2	18.9	110880	9	AC004878	AC004878 Homo sapi

ALIGNMENTS

RESULT 1  
AX127757  
LOCUS AX127757 416 bp DNA PAT 15-MAY-2001  
DEFINITION Sequence 10 from Patent WO0131042.  
ACCESSION AX127757  
VERSION AX127757.1 GI:14134404  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
synthetic construct.  
artificial sequence.  
REFERENCE  
AUTHORS 1 (bases 1 to 416)  
TITLE Weston,B. and de Beuckeleer,M.  
JOURNAL Male-sterile brassica plants and methods for producing same  
Patent: WO 0131042-A 10 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
1. 416  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="3' border flanking region of elite event MS-B2"



```

ACCESSION U66722
VERSION U66722.1 GI:1871363
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 110434)
AUTHORS BolDOG,F., Gemmill,R.M., West,J., Robinson,M., Robinson,L., Li,E.,
Roche,J., Todd,S., Waggoner,B., Lundstrom,R., Jacobson,J.,
Mullokandov,M.R., Klinger,H. and Drabkin,H.A.
Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B
Hum. Mol. Genet. 6 (2), 193-203 (1997)
97217778
TITLE 2 (bases 1 to 110434)
JOURNAL Robinson,H.A., BolDOG,F., West,J., Roche,J., Robinson,L.,
MEDLINE Robinson,M., Lundstrom,R., Li,E.F., Klinger,H. and Gemmill,R.M.
AUTHORS Direct Submission
TITLE Submitted (12-AUG-1996) Medical Oncology, University of Colorado
JOURNAL Health Sciences Center, 4200 E 9th Ave Box B171, Denver, CO 80262,
USA
COMMENT On Mar 8, 1997 this sequence version replaced gi:1669854.
FEATURES
Source
1..110434
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p14.2"
325..828
/rept_family="MER6"
repeat_region 920..1040
/rept_family="microsatellite"
repeat_region 933..991
/rept_family="HSAT1"
repeat_region 1287..1579
/rept_family="ALU"
repeat_region complement(4624..4730)
/rept_family="MIR"
LTR 7863..7887
/note="LTR8"
repeat_region 8276..8340
/rept_family="microsatellite"
LTR complement(9352..9405)
/note="LTR5"
repeat_region 10859..12771
/rept_family="Tigger1"
repeat_region complement(13165..13731)
/rept_family="LI"
repeat_region 13729..16089
/rept_family="LI"
repeat_region complement(17236..17352)
/rept_family="ALU"
repeat_region 18542..18838
/rept_family="ALU"
repeat_region complement(20464..20563)
/rept_family="MIR"
repeat_region complement(21303..21595)
/rept_family="ALU"
repeat_region 23140..23157
/rept_family="HSAT1"
repeat_region 24152..24196
/rept_family="MLTID"
repeat_region 24219..24274
/rept_family="MLTID"
repeat_region 25663..25913
/rept_family="ALU"
repeat_region 26593..26860
/rept_family="MER7"
repeat_region 26862..26936
/rept_family="MER7"
repeat_region complement(28463..28533)
/rept_family="MERS"
repeat_region complement(28903..29000)
/rept_family="LI"
/rept_family="MIR"
32445..32515
/note="LTR3"
repeat_region 33949..34268
/rept_family="LI"
repeat_region 34584..35022
/rept_family="LI"
LTR 35062..35099
/note="LTR6"
repeat_region complement(35225..35379)
/rept_family="MIR"
repeat_region 35771..35799
/rept_family="ALR"
misc_feature 36009..36120
/note="polypurine"
repeat_region complement(36138..36295)
/rept_family="ALU"
repeat_region complement(37577..37610)
/rept_family="XTR"
repeat_region 39771..39912
/rept_family="microsatellite"
43071..43274
/rept_family="microsatellite"
repeat_region complement(43199..43410)
/rept_family="ALU"
repeat_region complement(47356..47550)
/rept_family="MER20"
repeat_region complement(47798..47908)
/rept_family="TAR1"
repeat_region complement(48812..48924)
/rept_family="MERS"
repeat_region 50409..50541
/rept_family="MSTB"
repeat_region 50444..50541
/rept_family="MLTID"
repeat_region complement(52329..52447)
/rept_family="MIR"
repeat_region complement(53748..54028)
/rept_family="ALU"
repeat_region complement(56825..57121)
/rept_family="ALU"
repeat_region 58470..58533
/rept_family="MIR"
repeat_region 61841..62112
/rept_family="ALU"
repeat_region complement(62331..62604)
/rept_family="ALU"
repeat_region 64061..64195
/rept_family="LI"
misc_feature 64331..64468
/note="similar to EST R787000"
repeat_region complement(65817..66065)
/rept_family="LI"
repeat_region 66100..66588
/rept_family="MLTID"
repeat_region 66784..67036
/rept_family="LI"
repeat_region 67056..67843
/rept_family="LI"
repeat_region complement(68636..68913)
/rept_family="ALU"
LTR 69074..69099
/note="LTR6"
repeat_region 69253..69371
/rept_family="ALU"
repeat_region 71295..71324
/rept_family="XBR"
repeat_region complement(72381..72710)
/rept_family="ALU"
repeat_region complement(72953..73047)
/rept_family="LI"
repeat_region complement(74142..74787)
/rept_family="LI"

```







```

/note="assembly_fragment:01221
fragment_chain:1
clone_end:SP6
vector_side:left"
misc_feature
6553_-142126
/note="assembly_fragment:01911
fragment_chain:1
clone_end:T7
vector_side:right"
BASE COUNT 47396 a 26793 c 24991 g 42846 t 100 others
ORIGIN

Query Match 20.6%; Score 46; DB 2; Length 142126;
Best Local Similarity 54.0%; Pred. NO. 0.16;
Matches 94; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 10 ttataataattataataactgaaccattggtgccccctgctgctgttaccatgattt 69
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124601 TTATGCATACCTTAGATATATATATATATATATATATATATATATATATAT 124660

QY 70 ctcgcctactattgtatcagctgtatatataccgtataatgtacatatatttatgaac 129
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124661 ATATATATATATATATATATATATATATATATATATATATATATATATAT 124720

QY 130 atgattaatgctgtgagtttcttcctcgcgaagagtttcaatatgtaattggt 183
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124721 CTCTTAATACATGATCATCATACACCTATATCTATATACCTATATCTACAGTT 124774

RESULT 7
AC091684
LOCUS AC091684 81503 bp DNA HTG 17-MAY-2001
DEFINITION Homo sapiens chromosome UNK clone RP13-581E15, *** SEQUENCING IN
PROGRESS ***, 41 unordered pieces.
ACCESSION AC091684
VERSION AC091684.1 GI:14140336
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 81503)
Waterston,R.H.
Direct Submission
Submitted (17-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1406: contig of 1406 bp in length
* 1407 1506: gap of unknown length
* 1507 3416: contig of 1910 bp in length
* 3417 3516: gap of unknown length
* 3517 3161: contig of 2645 bp in length
* 3162 6261: gap of unknown length
* 6262 7972: contig of 1711 bp in length
*
8072: gap of unknown length
11094: contig of 3022 bp in length
11194: gap of unknown length
13957: contig of 2763 bp in length
14057: gap of unknown length
16589: contig of 2532 bp in length
16690: gap of unknown length
18911: contig of 2222 bp in length
21691: contig of 2680 bp in length
21791: gap of unknown length
24179: contig of 2288 bp in length
26194: contig of 2015 bp in length
26294: gap of unknown length
29162: contig of 2868 bp in length
29262: gap of unknown length
33740: contig of 4478 bp in length
33840: gap of unknown length
35244: contig of 1404 bp in length
35344: gap of unknown length
36397: contig of 1053 bp in length
36497: gap of unknown length
37724: contig of 1227 bp in length
37824: gap of unknown length
39046: contig of 1222 bp in length
39146: gap of unknown length
40876: contig of 1730 bp in length
40976: gap of unknown length
42169: contig of 1193 bp in length
42269: gap of unknown length
43531: contig of 1262 bp in length
43631: gap of unknown length
44992: contig of 1361 bp in length
45092: gap of unknown length
46384: contig of 1292 bp in length
46484: gap of unknown length
47767: contig of 1283 bp in length
47867: gap of unknown length
49395: contig of 1528 bp in length
49495: gap of unknown length
51364: contig of 1869 bp in length
51464: gap of unknown length
53294: contig of 1830 bp in length
53394: gap of unknown length
54985: contig of 1591 bp in length
55085: gap of unknown length
56463: contig of 1378 bp in length
56563: gap of unknown length
58670: contig of 2107 bp in length
58770: gap of unknown length
60099: contig of 1329 bp in length
60199: gap of unknown length
61632: contig of 1433 bp in length
61732: gap of unknown length
63819: contig of 2087 bp in length
63919: gap of unknown length
65179: contig of 1260 bp in length
65279: gap of unknown length
67578: contig of 2299 bp in length
67678: gap of unknown length
70151: contig of 2473 bp in length
70251: gap of unknown length
72332: contig of 1981 bp in length
72332: gap of unknown length
74165: contig of 1833 bp in length
74265: gap of unknown length
75644: contig of 1379 bp in length
75744: gap of unknown length
77594: contig of 1850 bp in length
77694: gap of unknown length
78888: contig of 1194 bp in length
78988: gap of unknown length

```



[illegible]





[illegible]









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:36 ; Search time 716.55 Seconds  
(without alignments)  
266.811 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416  
Perfect score: 223  
Sequence: 1 ctaccataattataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	100.0	416	22	Left (3') border f
2	189.8	85.1	415	22	Right (5') border
3	50.4	22.6	244	22	Oligonucleotide D1
4	50	22.4	936	22	Oligonucleotide D1
5	50	22.4	936	22	Oligonucleotide D1
6	50	22.4	936	22	Oligonucleotide D1
7	50	22.4	936	22	Oligonucleotide D1
8	50	22.4	936	22	Oligonucleotide D1
9	50	22.4	936	22	Oligonucleotide D1
10	50	22.4	936	22	Oligonucleotide D1
11	50	22.4	936	22	Oligonucleotide D2

12	50	22.4	936	22	AAF58262	Oligonucleotide D2
13	50	22.4	936	22	AAF58262	Oligonucleotide D2
14	50	22.4	938	22	AAF58255	Oligonucleotide D1
15	50	22.4	938	22	AAF58255	Oligonucleotide D1
16	47.4	21.3	244	22	AAF58238	Oligonucleotide D1
17	44	19.7	1153	21	AAC39525	Oligonucleotide D1
18	44	19.7	80240	20	AAV83940	Arabidopsis thalia
19	44	19.7	80240	20	AAV83939	NC-contig derived
20	40.8	18.3	1152	21	AAC46330	Arabidopsis thalia
21	38.8	17.4	513445	22	AAI61373	Soybean 318013 reg
22	38.2	17.1	1119	8	AAV71222	Sequence of soybean
23	37.4	16.8	32190	22	AAI62927	Human genomic DNA
24	37	16.6	470	22	AAH93301	Plasmodium falciparum
25	37	16.6	50000	21	AAA64139	Nucleotide sequenc
26	36.8	16.5	426	22	AAH93297	Arabidopsis thalia
27	36.8	16.5	15577	19	AAV35616	SHOX gene prelinin
28	36.8	16.5	32367	19	AAV35620	Human SHOX (short
29	36.6	16.4	893	22	AAH03844	Human CDNA clone (
30	36.6	16.4	1602	22	AAH15369	Human CDNA sequenc
31	36.4	16.3	2418	13	AAQ27886	P.falciparum GBP13
32	36.4	16.3	32249	22	AAI62932	Human genomic DNA
33	36.2	16.2	412	22	AAH93300	Plasmodium falciparum
34	35.8	16.1	916	22	AAH05774	Arabidopsis thalia
35	35.8	16.1	1558	21	AAH1567	N. meningitidis pa
36	35.8	16.1	335913	22	AAI61371	Soybean 240017 reg
37	35.8	16.1	335913	22	AAI61372	Soybean 240017 reg
38	35	15.7	41100	22	AAH10873	Human genomic DNA
39	35	15.7	101786	21	AAV22293	BAC containing rep
40	34.8	15.6	15418	21	AAH3785	Nucleotide sequenc
41	34.8	15.6	16875	22	AAH21613	Human hypocrerin r
42	34.8	15.6	16875	22	AAH21613	Human hypocrerin r
43	34.6	15.5	6038	21	AAZ43720	Human FSH-beta DNA
44	34.6	15.5	7622	21	AAZ43718	Human FSH-beta DNA
45	34.4	15.4	2451	20	AAV82439	Vascular endotheli

ALIGNMENTS

RESULT 1  
AAD06999 standard; DNA; 416 BP.  
ID AAD06999 standard; DNA; 416 BP.  
XX  
AC AAD06999;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE Left (3') border flanking region of elite event MS-B2.  
XX  
KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; ds.  
XX  
OS Chimeric - Agrobacterium sp.  
OS Chimeric - Brassica sp.  
FH Key Location/Qualifiers  
FT misc\_feature 1..193  
FT FT /\*tag= a  
FT FT /note= "Corresponds to T-DNA"  
FT FT misc\_feature 194..416  
FT FT /\*tag= b  
FT FT /note= "Corresponds to plant DNA"  
XX  
XX WO200131042-A2.  
XX  
XX 03-MAY-2001.  
XX  
XX 26-OCT-2000; 2000WO-EP10680.  
XX  
XX 29-OCT-1999; 99US-0430497.  
XX  
XX (AVET ) AVENTIS CROPS SCIENCE NV.  
XX



[illegible]





[illegible]



	XX	24-APR-2001 (first entry)	
	DT	Oligonucleotide D2007.	
	XX		
	DE	Electron-transfer group; ETW; mismatch; genotyping;	
	XX	gene expression; ss.	
KW	KW	Synthetic.	
XX	OS		
XX	PN	WO200107665-A2.	
XX	XX		
PD	PD	01-FEB-2001.	
XX	XX		
XX	PF	26-JUL-2000; 2000WO-US20476.	
PR	PR	26-JUL-1999; 99US-0145695.	
XX	PR	17-MAR-2000; 2000US-0190259.	
XX	PA	(CLIN-) CLINICAL MICRO SENSORS INC.	
XX	PI		
PT	PT	Umek RM;	
DR	DR	WPI; 2001-159728/16.	
XX	XX	Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface .	
PS	PS	Example 6; Page 128; 159pp; English.	
CC	CC	The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETW) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping,	
XX	XX	monitoring gene expression.	
SQ	SQ	Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;	
		Query Match            22.4%; Score 50; DB 22; Length 936;	
		Best Local Similarity     1.0%; Pred. No. 0.00033;	
		Matches      2; Conservative 141; Mismatches 61; Indels       0; Gaps	
QY	QY	6 ataattataatataactgaaaccacgtgcgccctcgtcgtttcacatgg 65 ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : :	
Db	Db	15 www..... ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: :	
QY	QY	66 atttcocgc tactatttg taca cgtgat atatacc gtaata acgtacatatatttat 125 ::: : ::: : :::   ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: :	
Db	Db	75 ..... ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: :	
QY	QY	126 gaacatgaattaagcttggagtgttgtctcatcccogaaggagtccaatagaatgtga 185 ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: :	
Db	Db	135 ..... ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: :	
QY	QY	186 agagtc caaac cc aaac catcaa 209 !  : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: :	
Db	Db	195 wgww..... ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: : ::: :	
		RESULT 13	
ID	ID	AAF58262/c	
XX	XX	AAF58262 standard; DNA; 936 BP.	
XX	XX	AAF58262:	
DT	DT	24-APR-2001 (first entry)	
DE	DE	Oligonucleotide D2007.	
XK	XK	Electron-transfer group; ETW; mismatch; genotyping;	



---

**THIS PAGE BLANK (USP 10)**







;; FILING DATE: 16-JUN-94  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/10814  
;; FILING DATE: 29-OCT-93  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: USSN 07/968,971  
;; FILING DATE: 30-OCT-92  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Elizabeth Lassen  
;; REGISTRATION NUMBER: 31,845  
;; NAME: Donna E. Scherer  
;; REGISTRATION NUMBER: 34,719  
;; NAME: Carl J. Schwedler  
;; REGISTRATION NUMBER: 36,924  
;; REFERENCE/DOCKET NUMBER: CGNE 111  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (916) 753-6313  
;; TELEFAX: (916) 753-1510  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1983 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA to mRNA  
US-08-383-736-3

Query Match 14.7%; Score 32.8; DB 1; Length 1983;  
Best Local Similarity 54.0%; Pred. No. 0.95;  
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 49 cctgctgtttacatggatttctccgctactattgtatcgtgtatataaccgtataa 108  
Db 1718 CGTGTGCTGCGCTGCAGCTGCTCCCAATATATATATATATATATATATATA 1599  
QY 109 tttacatatatttatgaacatgatttaattgtgtgagttgttctcatcgaagatt 168  
Db 1658 TATATATATATATATATATATATACATTTTGGCTATGCATTATTGTCTGGAGGAATCA 1599  
QY 169 tcaa 172  
Db 1598 TCAA 1595

RESULT 9  
US-08-460-898-3/c  
; Sequence 3, Application US/08460898  
; Patent No. 5850022  
; GENERAL INFORMATION:  
; APPLICANT: Dehesh, Katayoon  
; APPLICANT: Voelker, Toni Alois  
; APPLICANT: Hawkins, Deborah  
; APPLICANT: Davies, Huw Maelor  
; TITLE OF INVENTION: Production of Myristate in Plant Cells  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calgene, Inc.  
; STREET: 1920 Fifth Street  
; CITY: Davis  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 95616  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word 5.1(a)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,898  
; FILING DATE: 05-JUN-1995

;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/383,756  
;; FILING DATE: 02-FEB-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/261,695  
;; FILING DATE: 16-JUN-94  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/10814  
;; FILING DATE: 29-OCT-93  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: USSN 07/968,971  
;; FILING DATE: 30-OCT-92  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Elizabeth Lassen  
;; REGISTRATION NUMBER: 31,845  
;; NAME: Donna E. Scherer  
;; REGISTRATION NUMBER: 34,719  
;; NAME: Carl J. Schwedler  
;; REGISTRATION NUMBER: 36,924  
;; REFERENCE/DOCKET NUMBER: CGNE 111-1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (916) 753-6313  
;; TELEFAX: (916) 753-1510  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1983 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA to mRNA  
US-08-460-898-3

Query Match 14.7%; Score 32.8; DB 2; Length 1983;  
Best Local Similarity 54.0%; Pred. No. 0.95;  
Matches 67; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 49 cctgctgtttacatggatttctccgctactattgtatcgtgtatataaccgtataa 108  
Db 1718 CGTGTGCTGCGCTGCAGCTGCTCCCAATATATATATATATATATATATATA 1659  
QY 109 tttacatatatttatgaacatgatttaattgtgtgagttgttctcatcgaagatt 168  
Db 1658 TATATATATATATATATATATATACATTTTGGCTATGCATTATTGTCTGGAGGAATCA 1599  
QY 169 tcaa 172  
Db 1598 TCAA 1595

RESULT 10  
US-09-058-947A-1/c  
; Sequence 1, Application US/09058947A  
; Patent No. 6274790  
; GENERAL INFORMATION:  
; APPLICANT: Kunst et al.  
; TITLE OF INVENTION: Nucleic Acids Encoding Plant Enzyme  
; TITLE OF INVENTION: Involved in Very Long Chain Fatty Acid Synthesis  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klarquist Sparkman Campbell  
; ADDRESSEE: Leigh & Whinston, LLP  
; STREET: One World Trade Center, Suite  
; STREET: 1600, 121 S.W. Salmon Street  
; CITY: Portland  
; STATE: OR  
; COUNTRY: USA  
; ZIP: 97204-2988  
; COMPUTER READABLE FORM:



; MEDIUM TYPE: Disk, 3.5-inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Windows NT  
; SOFTWARE: Word97 & ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/058, 947A  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/043,831  
; FILING DATE: April 14, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: David J. Barp, Ph.D.  
; REGISTRATION NUMBER: 41,401  
; REFERENCE/DOCKET NUMBER: 5493-50032/DJE  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (503) 226-7391  
; TELEFAX: (503) 228-9446  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3722  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-09-058-947A-1

Query Match 14.7%; Score 32.8; DB 4; Length 3722;  
Best Local Similarity 56.2%; Pred. No. 1.1;  
Matches 82; Conservative 0; Mismatches 62; Indels 2; Gaps 1;  
  
Qy 61 catggatttcctactattgtatcacgtgtatatataccgtataatgtacatatatt 120  
Db 892 CGTAAATACAGAGAAATATAGGTACATATATATTCGGGTTTAAATGTCGATACATG 833  
  
Qy 121 tatatgaacatga--ttaatgcttgtgagttgtctcatccgttaagagtttcaatatgta 178  
Db 832 TATATCATACAGAGTTACATGTTTGTGTTTACATTCGCAATTAATGTTCTTTTTCG 773  
  
Qy 179 atgdtgaagagtcataaaaccccaaatc 204  
Db 772 ATGCTTAATCGTCAACACCGATATC 747

RESULT 11  
US-08-947-823-1  
; Sequence 1, Application US/08947823  
; Patent No. 6114605  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Valerie M.  
; APPLICANT: Kaloshian, Isouhli  
; APPLICANT: Yaghoobi, Jafar  
; APPLICANT: Bodeau, John  
; APPLICANT: Milligan, Stephen  
; TITLE OF INVENTION: Procedures and Materials for Confering  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/947,823  
; FILING DATE: 09-OCT-1997  
; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/18802  
; FILING DATE: 09-OCT-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/028,191  
; FILING DATE: 10-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-070210US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 51952 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-947-823-1

Query Match 14.5%; Score 32.4; DB 3; Length 51952;  
Best Local Similarity 47.5%; Pred. No. 2.6;  
Matches 96; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
  
Qy 5 cataataataataataataactgaacacgaagtgccccctgctgttacctg 64  
Db 13606 CATCTACGTATCTATAAAATTCACATGAACACACGCTGGAAGCGCATCTCAGAA 13665  
  
Qy 65 gatttcctccgactattgtatcacgtgtatatataccgtataatgtacatatattata 124  
Db 13666 AACTAGTGTA 13725  
  
Qy 125 tgaacatgattaaagctgtgagttgtctcatccgtaagaggtttcaatatgtatggtg 184  
Db 13726 TAT 13785  
  
Qy 185 aagagtcataaaaccccaaatcat 206  
Db 13786 ATCTGGTTAAAAAAGATAAT 13807

RESULT 12  
US-08-443-639-7/c  
; Sequence 7, Application US/08443639  
; Patent No. 5981843  
; GENERAL INFORMATION:  
; APPLICANT: Chappell, Joseph  
; APPLICANT: Yin, Shaohui  
; APPLICANT: Cornett, Catherine A.G.  
; TITLE OF INVENTION: Transcriptional Control Sequences and  
; TITLE OF INVENTION: Methods  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/443,639  
; FILING DATE: 18-MAY-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.

REGISTRATION NUMBER: 33878  
REFERENCE/DOCKET NUMBER: 69-94  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4254 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Nicotiana tabacum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(1217..1327, 1455..1718, 1806..2182,  
LOCATION: 2259  
LOCATION: ..2477, 2609..2747, 2903..3148, 3262..3558)  
US-08-443-639-7

Query Match 14.3%; Score 32; DB 2; Length 4254;  
Best Local Similarity 48.9%; Pred. No. 1.9;  
Matches 86; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 10 ttataattataattataactgaaccatggtgccccctgctgtttacatggattt 69  
Db 408 TTAITTTTCAAGAAAAAAGTTGCGTGGCGCGCGCATGTCACACACACACAC 349  
Qy 70 ctccgctactatctgtatcagtgatataaccgtataatgacatcatatttatgaac 129  
Db 348 ATACATTAT 289  
Qy 130 atgattaatgctgtgagttgtctcctccgtgaagagttccaatgtaagtga 185  
Db 288 ATATATTAATCTATAATTTTCTTTAACTTTTCACAAATTTTAGTGTATTTTCTGA 233

RESULT 13  
US-08-446-855A-1/c  
Sequence 1, Application US/08446855A  
Patent No. 5849573  
GENERAL INFORMATION:  
APPLICANT: Stewart, Thomas S  
APPLICANT: Flores, Maria V  
APPLICANT: O'Sullivan, William J  
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446.855A  
FILING DATE: 06-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mitchard, Leonard C  
REGISTRATION NUMBER: 29.009  
REFERENCE/DOCKET NUMBER: 47-80  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8920 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic  
US-08-446-855A-1

Query Match 14.3%; Score 32; DB 2; Length 8920;  
Best Local Similarity 54.2%; Pred. No. 2.2;  
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 77 actatttgatcagtgatataaccgtataatgacatataatttatgaacatgatta 136  
Db 8544 AATTA 8485  
Qy 137 atgcttgagtggtttctccatccgtgaagagtttccaatgtaagtgtgaagagtcacaaac 196  
Db 8484 TACTTTTCATGTTGTTATGGAATGAAAGTTAATAACATACATAAATAAGTAATGTTAAATC 8425

RESULT 14  
US-09-150-741-1/c  
Sequence 1, Application US/09150741  
Patent No. 6183996  
GENERAL INFORMATION:  
APPLICANT: Stewart et al.  
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
Patent No. 6183996  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/09/150,741  
EARLIER FILING DATE: 1998-09-10  
EARLIER APPLICATION NUMBER: PL6380  
EARLIER FILING DATE: 1992-12-16  
EARLIER APPLICATION NUMBER: AU93/00617  
EARLIER FILING DATE: 1993-12-02  
EARLIER APPLICATION NUMBER: 08/446.855  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 8920  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
US-09-150-741-1

Query Match 14.3%; Score 32; DB 4; Length 8920;  
Best Local Similarity 54.2%; Pred. No. 2.2;  
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 77 actatttgatcagtgatataaccgtataatgacatataatttatgaacatgatta 136  
Db 8544 AATTA 8485  
Qy 137 atgcttgagtggtttctccatccgtgaagagtttccaatgtaagtgtgaagagtcacaaac 196  
Db 8484 TACTTTTCATGTTGTTATGGAATGAAAGTTAATAACATACATAAATAAGTAATGTTAAATC 8425

RESULT 15  
US-08-282-581-4  
Sequence 4, Application US/08282581  
Patent No. 5670349  
GENERAL INFORMATION:  
APPLICANT: Cramer, Carole L.  
APPLICANT: Weissenborn, Deborah L.  
TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM AND



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:10 ; Search time 8261.74 Seconds  
(without alignments)  
290.049 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416

Perfect score: 223

Sequence: 1 ctaccataattataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	45.6	20.4	507	10	AA605352 30388 Lam
C 2	44.4	19.9	368	10	AU088482
C 3	43.6	19.6	465	11	BE948306
C 4	42	18.8	1024	13	CNS030D0
C 5	41.8	18.7	487	10	AU087926
C 6	41.8	18.7	500	10	AU088152
C 7	41.6	18.7	527	13	AQ172252
C 8	41.6	18.7	599	13	AQ273686
C 9	41.6	18.7	711	13	AZ119444
C 10	41.6	18.7	789	13	AZ126726
C 11	40.6	18.2	571	13	AQ797772
C 12	40.6	18.2	605	13	AZ996260

C 13	40.4	18.1	583	13	AQ774291
C 14	40.4	18.1	595	10	AA394343
C 15	40.4	18.1	845	13	AQ745537
C 16	40.2	18.0	525	11	BF460992
C 17	39.8	17.8	612	13	AZ409529
C 18	39.8	17.8	836	13	AZ908564
C 19	39.6	17.8	456	11	BG836222
C 20	39.4	17.7	458	13	BH040914
C 21	39.4	17.7	765	13	AZ705201
C 22	39.4	17.7	939	13	CNS0403P
C 23	39.2	17.6	363	13	AZ598385
C 24	39.2	17.6	1020	13	CNS03AUS
C 25	39.2	17.6	1101	13	CNS00DGO
C 26	39	17.5	309	13	AQ542682
C 27	39	17.5	511	11	BG633595
C 28	39	17.5	610	11	BF296565
C 29	39	17.5	764	13	AQ915359
C 30	39	17.5	834	10	AW941790
C 31	38.6	17.3	374	13	BH126294
C 32	38.6	17.3	453	13	AQ818698
C 33	38.6	17.3	596	13	BH126267
C 34	38.4	17.2	500	10	AU087788
C 35	38.4	17.2	536	13	AZ013131
C 36	38.4	17.2	563	13	AQ561057
C 37	38.2	17.1	192	10	AA547848
C 38	38.2	17.1	543	13	AZ060650
C 39	38.2	17.1	604	13	AZ347990
C 40	38	17.0	514	13	AQ141523
C 41	38	17.0	541	13	AQ640886
C 42	38	17.0	609	13	AQ367613
C 43	37.8	17.0	324	13	BH126635
C 44	37.8	17.0	412	10	A1991410
C 45	37.8	17.0	416	13	AZ869448

ALIGNMENTS

RESULT 1  
AA605352/c  
LOCUS AA605352 507 bp mRNA EST 29-SEP-1997  
DEFINITION 30388 Lambda-PRL2 Arabidopsis thaliana cDNA clone 269H877, mRNA sequence.  
ACCESSION AA605352  
VERSION AA605352.1 GI:2445880  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
REFERENCE 1 (bases 1 to 507)  
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel,E. and Somerville,C.  
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones  
JOURNAL Plant Physiol. 106, 1241-1255 (1994)  
MEDLINE 95148729  
COMMENT Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313tcn@bm.cl.msu.edu  
Seq primer: T7 dye primer.  
Location/Qualifiers  
1. 507  
/organism="Arabidopsis thaliana"  
/strain="var Columbia"  
/db\_xref="taxon:3702"



(cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtraced libraries were generated in this process: NIH\_BMAP\_M\_S4, NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1, NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtraced library (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life technologies) to generate the NIH\_BMAP\_M\_S4 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_TISSUE=cerebellum  
TAG\_SEQ=GACTC"

BASE COUNT. 174 a 67 c 60 g 164 t  
ORIGIN

Query Match 19.6%; Score 43.6; DB 11; Length 465;  
Best Local Similarity 53.5%; Pred. No. 0.69;  
Matches 91; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
  
Qy 8 aattataataattataataactgaaccatggtgccccctgctgtttacatgat 67  
Db 212 ACTTGGCATTAGACCCCTAAATTTCTGTCACAACTGATATATATATATATATATAT 271  
  
Qy 68 ttctccgtactattgtatcgtgtatatataccgtataatgtacatatatttatga 127  
Db 272 AT 331  
  
Qy 128 acatgattaactgtgtgagttgtctcatccgtgaagagtttcaatgt 177  
Db 332 ACATATATAATTATAAAGGTGTTTCGAGAAATAACACTTACAAACTGT 381

RESULT 4  
LOCUS CNS030D0 1024 bp DNA GSS 15-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 184002 of library G from Tetraodon nigroviridis, genomic survey sequence.  
ACCESSION AL222093.1 GI:7880912  
VERSION GSS: genome survey sequence.  
KEYWORDS Tetraodon nigroviridis.  
SOURCE Tetraodon nigroviridis  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
REFERENCE 1 (bases 1 to 1024)  
AUTHORS Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.  
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1024)  
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.  
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1024)  
AUTHORS Genoscope.  
TITLE Direct Submission  
COMMENT Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.  
FEATURES  
source  
1. 1024  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="184002"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG184BH01LPI-end : T7"  
BASE COUNT 346 a 70 c 174 g 333 t 101 Others  
ORIGIN  
  
Query Match 18.8%; Score 42; DB 13; Length 1024;  
Best Local Similarity 41.1%; Pred. No. 1.7;  
Matches 88; Conservative 23; Mismatches 103; Indels 0; Gaps 0;  
  
Qy 6 ataattataattataataactgaaccatggtgccccctgctgtttacatgg 65  
Db 463 ATACAT 522  
  
Qy 66 atttcgcgtactattgtatcgtgtatatataccgtataatgtacatatatttat 125  
Db 523 ATTTTMTTTTAAATATATATATTTTAAAWATATATATATATATATATATATATATAT 582  
  
Qy 126 gaacatgtaactgtgtgagttgtctcatccgtgaagagtttcaatgtgaatgg 185  
Db 593 ATATATATATATATATATATTTTWTWTWTWTATATATATATATATATATATATATAT 642  
  
Qy 186 agagtcaaaaccccaatcatgacaccccaact 219  
Db 643 TTWTATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 676  
  
RESULT 5  
LOCUS AU087926/c 487 bp mRNA EST 27-JAN-2001  
DEFINITION AU087926 Sugano Malaria cDNA library Plasmodium falciparum CDNA clone XPN6146, mRNA sequence.  
ACCESSION AU087926.1 GI:12390067  
VERSION AU087926  
KEYWORDS EST.  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
REFERENCE 1 (bases 1 to 487)  
AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.  
TITLE FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum  
JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)  
MEDLINE 20574754  
COMMENT Contact: Junichi Watanabe  
Institute of Medical Science  
The University of Tokyo, Department of Parasitology  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Tel: 81-3-5449-5378  
Fax: 81-3-5449-5410  
Email: jwatanabe@med.s.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).  
FEATURES  
source  
1. 487  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"

```

/clone="XPFn6146"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage" 6 others
BASE COUNT      213 a      23 c      23 g      211 t
ORIGIN

Query Match      18.7%; Score 41.8; DB 10; Length 487;
Best Local Similarity 52.4%; Pred. No. 1.9;
Matches 88; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 8 aattataattataataataatactgaacacatggtgccccctgctgttacctgagt 67
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 AATATATATGAATAATATATGAAATATATATGATGATTTTCTTCATGCTGAT 109
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 68 ttctcgcgtactattgtatcagtggtatataacgtataatgacatatatttatga 127
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 TTGTATTANCCATATATATATATATATATATATGTTAATTAATAAATGTAGAAAC 49
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 128 acatgattcaatgctgtgaggtgtctcatccgtaagaggtttcaatat 175
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 GATTTTATTTTATTTATATATATATAAATCCGNAATAATTNCAAAAT 1

RESULT 6
LOCUS      AU088152      500 bp      mRNA      27-JAN-2001
DEFINITION AU088152 Sugano Malaria cDNA library Plasmodium falciparum cDNA
clone XPFn6606 similar to Plasmodium falciparum microsatellite Ta61
sequence, mRNA sequence.
ACCESSION AU088152
VERSION AU088152.1 GI:12390293
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 500)
Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
FULL-malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum
Nucleic Acids Res. 29 (1), 70-71 (2001).
20574754
Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanabe@manage.ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
source
1..500
Location/Qualifiers
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone="XPFn6606"
/dev_stage="erythrocytic stage"
BASE COUNT      224 a      37 c      38 g      182 t      19 others
ORIGIN

Query Match      18.7%; Score 41.8; DB 10; Length 500;
Best Local Similarity 55.5%; Pred. No. 1.9;
Matches 76; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 83 tgtatcagtggtatataacgtataatgtacatatattttatgacatgattatgctt 142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 TATTATNCTAAATATATATACATATATATATATATATATATATATATATATGTTTA 122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 gtaggtgttctcatccgtgaaggtttccaattgaatggtgaagcgaacccaaaaa 202
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 TTNATTTTAAATATATAAATATATATATATATATATATGTAAGAANGAAAGAAAA 182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 203 tcatgaacacccaaact 219
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 AAATTAAANGCAAATTT 199
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
LOCUS      AQ172252/c      527 bp      DNA      GSS      17-OCT-1998
DEFINITION HS_3179_B2_B11_T7 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3179 Col=22 Row=D, DNA sequence.
ACCESSION AQ172252
VERSION AQ172252.1 GI:3569619
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc.Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3179 row: D column: 22
Class: BAC ends
High quality sequence stop: 527.
FEATURES
source
1..527
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3179 Col=22 Row=D"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      195 a      56 c      55 g      218 t      3 others
ORIGIN

Query Match      18.7%; Score 41.6; DB 13; Length 527;
Best Local Similarity 70.0%; Pred. No. 2.1;
Matches 56; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 57 ttatcatggttctccgtactatttgtatcagtggtatataacgtataatgataata 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 TATATATAGATATATATAGATATATATATATATATATATATATATATATATATATA 344
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 117 tatttatgacatgatta 136
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 TATTATATATAAATAATAA 324
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
LOCUS      AQ273686      599 bp      DNA      GSS      03-NOV-1998
DEFINITION nbxb0030P21f CUGI Rice BAC Library Oryza sativa genomic clone
nbxb0030P21f, DNA sequence.
ACCESSION AQ273686
VERSION AQ273686.1 GI:3827001
KEYWORDS GSS.

```





```

ACCESSION   AZ126726
VERSION     AZ126726.1  GI:8201718
KEYWORDS    GSS.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
            1 (bases 1 to 789)
REFERENCE   Wing,R.A. and Dean,R.A.
            A BAC End Sequencing Framework to Sequence the Rice Genome
            Unpublished (1998)
TITLE       Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
COMMENT     Seq primer: GTAAACGACGCGCAGTG
            Class: BAC ends
            High quality sequence start: 21
            High quality sequence stop: 683.
            Location/Qualifiers
                1..789
                /organism="Oryza sativa"
                /strain="Japonica"
                /cultivar="Nipponbare"
                /db_xref="taxon:4530"
                /clone="OSJNB0079C19f"
                /clone_lib="CUGI Rice BAC Library (EcoRI)"
                /tissue_type="Leaf"
                /lab_host="E. coli DH10B"
                /note="vector: pBACIndigo; Site1: EcoRI; Site2: EcoRI;
            Rice is the most important food crop in the world. Half of
            the world population, especially those inhabiting highly
            populated areas of the humid tropics and subtropics, rely
            on rice as their primary source of carbohydrate.
            Monocytodonous rice is a diploid plant (2n=24) with a
            haploid genome equivalent of 431 Mbp (Arumuganathan and
            Earle, 1991). The relatively small genome of rice, three
            times larger than that of Arabidopsis, makes it suitable
            for genomic studies. In order to facilitate positional
            cloning, physical mapping and genome sequencing of rice,
            we have constructed a BAC library from Oryza sativa.
            Nipponbare variety using EcoRI as the cloning enzyme. The
            library contains 55,296 clones with an average insert size
            of 121 Kb providing approximately 15 haploid genome
            equivalents. The deep coverage allows the isolation a
            particular sequence with a probability of 99.9 %. Three
            high density filters, each containing 18,432 clones
            (doubly spotted), represent the whole library for colony
            screening and can be requested from the Clemson University
            BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT   292 a 124 c 111 g 253 t
ORIGIN
Query Match      18.7%; Score 41.6; DB 13; Length 789;
Best Local Similarity 50.9%; Pred. No. 2.1;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY  9 attataataataataactgaaccatggcgccctgctgttcacatgatt 68
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  745 ATTATATATATATATATATAAATATATATTTTATTTATTTTANATNTIN 686
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  69 tctccgctactattgtacatgtatatataccggtatattgtacatatattatgaa 128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  685 NATTANNNTATTTTATATATATATATATATATATATATATATATATATA 626
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  129 catgataatgcctgtgagttgttcacccgtaagagtttcaatgt 177
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  625 TATATATCTTCATTCAGCATGTTTCACATACAGATAGCTTTAGGTAT 577
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 11
AQ797772/c
LOCUS     AQ797772      571 bp      DNA      GSS      09-AUG-1999
DEFINITION HS_2161_E2_E10_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2161 Col=20 Row=J, DNA sequence.
ACCESSION   AQ797772
VERSION     AQ797772.1  GI:5715030
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 571)
REFERENCE   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
            99380589
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2161 row: J column: 20
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 571.
            Location/Qualifiers
                1..571
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate=2161 Col=20 Row=J"
                /clone_lib="CIT Approved Human Genomic Sperm Library D"
                /sex="male"
                /note="organ: sperm; Vector: pBelOBAC11; BAC Clones in
            E-Coli DH10B"
BASE COUNT   197 a 48 c 83 g 241 t
ORIGIN
Query Match      18.2%; Score 40.6; DB 13; Length 571;
Best Local Similarity 58.8%; Pred. No. 3.7;
Matches 70; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY  60 acatggatttccgctactattgtatcgtgtatatataccgtataatgacatatat 119
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  192 ACATATATATATACACATATATATGTATATACATATATATATATATATATATAT 133
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  120 ttatatgaacatgattaatgcttggtgagttgtcttcacgcgaagttccaatgta 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  132 ATATATACACATATATATGTATATACATATATATATATATATATATATATATAT 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 12
AZ996260/c
LOCUS     AZ996260      605 bp      DNA      GSS      27-APR-2001
DEFINITION 2M0282F21F Mouse 10kb plasmid UUGC2M library Mus musculus genomic clone UUGC2M0282F21 F, DNA sequence.
ACCESSION   AZ996260
VERSION     AZ996260.1  GI:13867487
KEYWORDS    GSS.
SOURCE      house mouse.
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:02:29 ; Search time 2331.3 Seconds  
(without alignments)  
162.757 Million cell updates/sec

Title: US-09-698-903B-11

Perfect score: 23  
Sequence: 1 gaaatccatgtaaacgacgagg 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_em.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	DB ID	Description
1	23	100.0	23	6	AX127758 Sequence
2	23	100.0	415	6	AX127755 Sequence
3	23	100.0	416	6	AX127757 Sequence
4	18.8	81.7	41760	2	AC090679 Homo sapi
5	18.8	81.7	89350	8	AF112H17 Arabidops
6	18.8	81.7	156515	2	AC022947 Homo sapi
7	18.8	81.7	199577	8	ATCHRIV57 Arabidops
8	18.8	81.7	208684	2	AC072048 Mus muscu
9	18.4	80.0	158524	2	AC046148 Mus muscu
10	18.4	80.0	172650	2	AC040920 Homo sapi
11	18.2	79.1	535	5	AF118429 Falco per
12	18.2	79.1	146285	9	AC005083 Homo sapi
13	18.2	79.1	183634	2	AC023167 Mus muscu
14	18.2	79.1	228809	2	AC021443 Homo sapi
15	17.8	77.4	23486	8	AC007294 Arabidops
16	17.8	77.4	26275	3	U29377 Caenorhabdi
17	17.8	77.4	94835	9	HSBJ316D7 Human DNA
18	17.8	77.4	139959	33	AC013597 Homo sapi
19	17.8	77.4	146746	9	HS03222 Human DNA s
20	17.8	77.4	149755	2	AC092795 Homo sapi
21	17.8	77.4	168749	2	AC008573 Homo sapi
22	17.8	77.4	177703	9	AC019046 Homo sapi
23	17.8	77.4	179665	2	AC010246 Homo sapi
24	17.8	77.4	188537	2	AC090071 Homo sapi
25	17.8	77.4	213890	2	AC013461 Homo sapi
26	17.8	77.4	253038	2	AC008930 Homo sapi
27	17.4	75.7	25042	9	HS380A1 Human DNA s
28	17.4	75.7	89939	2	AC022667 Homo sapi
29	17.4	75.7	106392	9	AC007316 Homo sapi
30	17.4	75.7	113033	9	AC005070 Homo sapi
31	17.4	75.7	118904	9	HS238G2 Human DNA
32	17.4	75.7	134506	9	HS357I16 Homo sapi
33	17.4	75.7	136497	2	HSBJ448H8 Homo sapi
34	17.4	75.7	155316	2	AC078902 Homo sapi
35	17.4	75.7	155932	1	CJ11168X6 Campyloba
36	17.4	75.7	156014	5	AF003796 Gallus ga
37	17.4	75.7	171368	9	HSU95738 Human chrom
38	17.4	75.7	171868	9	AL359697 Human DNA
39	17.4	75.7	194873	2	AL592440 Homo sapi
40	17.4	75.7	256073	9	AE006464 Homo sapi
41	17.2	74.8	466	11	HSPE53B06 H. sapiens
42	17.2	74.8	522	5	RCABCRYST X87114 R. catesbeia
43	17.2	74.8	985	3	AB002187 Epiliachna
44	17.2	74.8	985	3	AB002188 Epiliachna
45	17.2	74.8	985	3	AB002189 Epiliachna

ALIGNMENTS

RESULT	1				
AX127758	AX127758	Sequence 11 from Patent WO0131042.	23 bp	DNA	15-MAY-2001
LOCUS	AX127758	Sequence 11 from Patent WO0131042.			
DEFINITION	AX127758				
ACCESSION	AX127758				
VERSION	AX127758.1	GI:14134405			
KEYWORDS					
SOURCE					
ORGANISM					
		synthetic construct.			
		synthetic construct			
		artificial sequence.			
REFERENCE		1 (bases 1 to 23)			
AUTHORS		Weston,B. and de Beuckeleer,M.			
TITLE		Male-sterile brassica plants and methods for producing same			
JOURNAL		Patent: WO 0131042-A 11 03-MAY-2001;			
		Aventis CropScience N.V. (BE)			
FEATURES		location/Qualifiers			
source		1..23			
		/organism="synthetic construct"			
		/db_xref="taxon:32630"			
		/note="primer MDB371"			

BASE COUNT 9 a 4 c 7 g 3 t  
ORIGIN

Query Match 100.0%; Score 23; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.26;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcagg 23  
|||||  
Db 1 GAAATCCATGTAAGCAGCAGG 23

## RESULT 2

AX127755  
LOCUS AX127755 415 bp DNA PAT 15-MAY-2001  
DEFINITION Sequence 8 from Patent WO0131042.  
ACCESSION AX127755  
VERSION AX127755.1 GI:14134402  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 415).  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 8 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source 1. .415  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="5' border flanking region of elite event MS-B2"  
misc\_feature 1. .234  
/note="plant DNA"  
misc\_feature 235. .415  
/note="T-DNA"

BASE COUNT 154 a 55 c 70 g 136 t  
ORIGIN

Query Match 100.0%; Score 23; DB 6; Length 415;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcagg 23  
|||||  
Db 152 GAAATCCATGTAAGCAGCAGG 174

## RESULT 3

AX127757/c  
LOCUS AX127757 416 bp DNA PAT 15-MAY-2001  
DEFINITION Sequence 10 from Patent WO0131042.  
ACCESSION AX127757  
VERSION AX127757.1 GI:14134404  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 416).  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source 1. .416  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="5' border flanking region of elite event MS-B2"  
misc\_feature 1. .193  
/note="T-DNA"  
misc\_feature 194. .416

BASE COUNT 137 a 72 c 54 g 152 t 1 others  
ORIGIN

Query Match 100.0%; Score 23; DB 6; Length 416;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcagg 23  
|||||  
Db 263 GAAATCCATGTAAGCAGCAGG 241

## RESULT 4

AC090679/c  
LOCUS AC090679 41760 bp DNA HTG 27-JUN-2001  
DEFINITION Homo sapiens chromosome 12q clone RP11-8003, \*\*\* SEQUENCING IN PROGRESS \*\*\*; 14 unordered pieces.  
AC090679  
VERSION AC090679.3 GI:14190581  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 41760)

## AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbarella,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Bathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lousegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

## TITLE

Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 41760)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One



exon /gene="T12H17.20"  
/number=1  
2844..2983  
/gene="T12H17.20"  
trRNA /number=2  
3327..3400  
/note="Asn-tRNA\_(GTT)"  
gene 4696..5489  
/gene="T12H17.30"  
CDS join(4696..5035,5407..5489)  
/gene="T12H17.30"  
/note="contains EST gb:N38564"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAA16549.1"  
/db\_xref="GI:2827541"  
/db\_xref="SPTREMBL:O49645"  
/translation="MAYTNKVTISAATVATMMLFLAVITVDAQSMPPMPKPNPYCALAD  
LPNIYQICYNLDITPSECCNDLKSSSTIQVNCICDNFLAHPNSGNISQARYDLVNS  
ACGVADKFAACKGGDASGGSTNKIAASNVLLGLVASLFF"  
exon /gene="T12H17.30"  
/number=1  
5036..5406  
/gene="T12H17.30"  
intron /number=1  
5407..5489  
/gene="T12H17.30"  
CDS join(6595..6940,7035..7105)  
/gene="T12H17.40"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAA16550.1"  
/db\_xref="GI:2827542"  
/db\_xref="SPTREMBL:O49646"  
/translation="MAVAVATVFLAVIAPQWETTKPPRESDTSDTSGTSGDRRT  
MCPLSGIVQNCYATLNAFPKCKDLKTAKREVTCLCNNVIAHPDPLYTNTNQV  
6595..6940  
/gene="T12H17.40"  
/number=1  
6595..7105  
gene /gene="T12H17.40"  
intron 6941..7034  
/gene="T12H17.40"  
exon /number=1  
7035..7105  
/gene="T12H17.40"  
exon /number=2  
complement(8019..9209)  
/gene="T12H17.50"  
/number=1  
8019..9209  
gene /gene="T12H17.50"  
gene complement(8019..9209)  
/gene="T12H17.50"  
complement(8019..9209)  
CDS /gene="T12H17.50"  
/note="similarity to predicted protein, Arabidopsis  
thaliana"  
/codon\_start=1  
/product="predicted protein"  
/protein\_id="CAA16551.1"  
/db\_xref="GI:2827543"  
/db\_xref="SPTREMBL:O49647"  
/translation="MEKNQNPNTWSDLPDLLNLVFKRLSFANFRQAKSCSSWYSAS  
KQSPKQIPIWMLFPKDKNNKNSCTIFFNPEDKQLYQTODLGVFAKSKYCLATY  
GSLWLDQDSKYNLYILNPFYKIGLPAIESQOVGMKVQDTIDDDFLTDDHNVKYL  
FKGNVTVPVFNIDEKTKYIALWGLGYWCVVYAKNGDKLWNLPEIILDSIDMWYK  
DHKLYSFYSNLFTILDFSGEIPRKAFCQCFMHVYRSPLSPRSQIWSNCVVAETKL  
VTVTGVDLLVRLHWSRIQSENKYKFSYSGFLDKYELASIGDEAMLLDITVL  
ANEVEGLHRTIYFSDSHDITTKDLFLNFETREMEPLHKFDKCSLLELSARWFLPS

FSHT"  
9990..13709  
/gene="T12H17.60"  
join(9990..10023,10239..10535,10564..10664,10910..11292,  
11755..11809,11898..12053,12143..12311,12391..12455,  
12534..12674,12769..12885,13024..13096,13244..13380,  
13468..13525,13612..13709)  
/gene="T12H17.60"  
/note="similarity to p48, Homo sapiens, PATCHX:G904032  
contains EST gb:Ag605571, H76919, H76638, R30271, Z25650"  
/codon\_start=1  
/product="HSP associated protein like"  
/protein\_id="CAA16552.1"  
/db\_xref="GI:2827544"  
/db\_xref="SPTREMBL:O49648"  
/translation="HRTWVLAESVAIAPLLAEPQSMFPEMTPECATVMPDLLEK  
CRATGSVTTECCDILKTSATQVTCLDNYIANPVSNTITGYSKAITTKGVDFK  
YSCDGTSKVISNKGAFMYINGFWCVVKEKKKEAAVAMGVTLGLINGLIVAH  
GKTRQNKPVCKLIILFQFHGFCISIRFSANKKRSKSTGDAYPFRKEKYLEQKPY  
LPFVQILEIETRLTKTKTRLLTRPKPTLFLSLSRSEADFQKQGFNEAYLGAK  
IPTGVHEEDKTPRSFVVEESDDMDTEEVKPKVEEEDEIIEHLTRAITLNPISA  
PNDPPQKMGDSSVEVDENRRAAQEAQKAMEALSEGFEDEIEHLTRAITLNPISA  
INYGNSAVYIKLKPNAIRDAANALEINPDSAKGYKSRGMARMLGEAAKDLH  
LASTIDYDEEISAVLKKVEPNNAHLEEHRRYDRLKREDEKKAERDLRRRAQA  
YDKAKKEQSSSSSSGGGPGMGGMGPAGMGGMGPAGMGGMGPAGMGGMGP  
GGMDFSKILNDPELMTAFSDPEVMAALQDVNMKNPANLAKHQANPKVAVIAKMGKFA  
GPQ"  
9990..10023  
/gene="T12H17.60"  
/number=1  
10024..10238  
/gene="T12H17.60"  
exon /number=1  
10239..10535  
/gene="T12H17.60"  
intron /number=2  
10536..10563  
/gene="T12H17.60"  
exon /number=2  
10564..10664  
/gene="T12H17.60"  
intron /number=3  
10665..10909  
/gene="T12H17.60"  
exon /number=3  
10910..11292  
/gene="T12H17.60"  
intron /number=4  
11293..11754  
/gene="T12H17.60"  
exon /number=4  
11755..11809  
/gene="T12H17.60"  
intron /number=5  
11810..11897  
/gene="T12H17.60"  
exon /number=5  
11898..12053  
/gene="T12H17.60"  
intron /number=6  
12054..12142

Query Match 81.7% Score 18.8 DB 8 Length 89350;  
Best Local Similarity 90.9% Pred. No. 95;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 aaatcccatgtaagcagcaggg 23  
|||||  
Db 15674 AAATCCCATGTAAGCAGCTAAGG 15653  
RESULT 6  
AC022947/c



**LOCUS** AC022947 156515 bp DNA HTG 12-MAR-2000  
**DEFINITION** Homo sapiens clone RP11-8003, WORKING DRAFT SEQUENCE, 13 unordered pieces.  
**ACCESSION** AC022947  
**VERSION** AC022947.2 GI:7229879  
**KEYWORDS** HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 156515)  
Birren, B., Linton, J., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-8003  
Unpublished  
2 (bases 1 to 156515)  
Birren, B., Linton, J., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Castle, A., Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 12, 2000 this sequence version replaced gi:6921807.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L6397  
Center clone name: 80\_O\_3  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 147422 bases at least Q40  
Consensus quality: 151928 bases at least Q30  
Consensus quality: 153939 bases at least Q20  
Insert size: 150000; agarose-fp  
Insert size: 153115; sum-of-contigs  
Quality coverage: 4.8 in Q20 bases; agarose-fp  
Quality coverage: 4.6 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence \* as soon as it is available and the accession number will be preserved.  
1 1026: contig of 1026 bp in length  
\* 1027 1126: gap of 100 bp  
\* 1127 5787: contig of 4661 bp in length  
\* 5788 5887: gap of 100 bp  
\* 5888 11595: contig of 5708 bp in length  
\* 11596 11695: gap of 100 bp

\* 11696 14431: contig of 2736 bp in length.  
\* 14432 14531: gap of 100 bp  
\* 14532 20088: contig of 5557 bp in length  
\* 20089 20188: gap of 100 bp  
\* 20189 29619: contig of 9431 bp in length  
\* 29620 29719: gap of 100 bp  
\* 29720 38087: contig of 8368 bp in length  
\* 38088 38187: gap of 100 bp  
\* 38188 50251: contig of 12064 bp in length  
\* 50252 50351: gap of 100 bp  
\* 50352 60663: contig of 10312 bp in length  
\* 60664 60763: gap of 100 bp  
\* 60764 75537: contig of 14774 bp in length  
\* 75538 75637: gap of 100 bp  
\* 75638 93139: contig of 17502 bp in length  
\* 93140 93239: gap of 100 bp  
\* 93240 114372: contig of 21133 bp in length  
\* 114373 114472: gap of 100 bp  
\* 114473 156515: contig of 42043 bp in length.  
FEATURES  
Location/Qualifiers  
source  
1..156515  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="RPC1-11 Human Male BAC"  
/clone="RP11-8003"  
1..1026  
/note="assembly\_fragment"  
misc\_feature  
1127..5787  
/note="assembly\_fragment"  
misc\_feature  
5888..11595  
/note="assembly\_fragment"  
misc\_feature  
11696..14431  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
14532..20088  
/note="assembly\_fragment"  
20189..29619  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:left  
29720..38087  
/note="assembly\_fragment"  
38188..50251  
/note="assembly\_fragment"  
50352..60663  
/note="assembly\_fragment"  
60764..75537  
/note="assembly\_fragment"  
75638..93139  
/note="assembly\_fragment"  
93240..114372  
/note="assembly\_fragment"  
114473..156515  
/note="assembly\_fragment"  
BASE COUNT 49638 a 27072 c 27170 g 51430 t 1205 others  
ORIGIN  
Query Match 81.7%; Score 18.8; DB 2; Length 156515;  
Best Local Similarity 90.9%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 gaaatcatgtataagcagcagg 22  
|||||  
Db 76912 GAATTCATGTAAAGCTGCAGG 76891  
RESULT 7  
ATCHRIV57/c  
LOCUS ATCHRIV57 199577 bp DNA 16-MAR-2000  
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 57.  
ACCESSION AL161557

VERSION	AL161557.2	GI:7269071	
KEYWORDS	thale cress.		
SOURCE	Arabidopsis thaliana		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 64578 to 155510)		
JOURNAL	Wedler, H., Wambutt, R., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.		
REFERENCE	Unpublished		
AUTHORS	2 (bases 141638 to 141797)		
JOURNAL	Volckaert, G., Grymonprez, B., Voet, M., Robben, J., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.		
REFERENCE	Unpublished		
AUTHORS	3 (bases 143670 to 199577)		
JOURNAL	Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.		
REFERENCE	Unpublished		
AUTHORS	4 (bases 48809 to 64829)		
JOURNAL	Robben, J., Grymonprez, B., Volckaert, G., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.		
REFERENCE	Unpublished		
AUTHORS	5 (bases 1 to 199577)		
JOURNAL	EU Arabidopsis sequencing, project.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de		
JOURNAL	Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@ebc.ac.uk		
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> this fragment has an overlap with ATCHRIV56 at the 5' end and an overlap with ATCHRIV58 at the 3' end.		
FEATURES	Location/Qualifiers		
source	1..199577		
	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="4"		
	6674..9057		
gene	/gene="AT4g22260"		
exon	complement(6674..6882)		
	/gene="AT4g22260"		
	/number=1		
gene	complement(join(6674..6882,7268..7339,7507..7575,7667..7775,7863..7978,8361..8454,8588..8777,8909..9057))		
	/gene="AT4g22260"		
CDS	complement(join(6674..6882,7268..7339,7507..7575,7667..7775,7863..7978,8361..8454,8588..8777,8909..9057))		
	/gene="AT4g22260"		
	/note="similarity to alternative oxidase, Mangifera indica, PIR2:S45035		
	contains EST gb:T42793, AI994896.1, AI997645.1, R30022, AA395166"		
	/codon_start=1		
	/product="putative protein"		
	/protein_id="CAB79181.1"		
	/db_xref="GI:7269072"		
	/translation="MAAISGISGTLTISRPLVTLRRRAVSYSSSHRLHLPLSS RRLLRNHRVQATILQDDEKVVVEESFKAETGTGTEPLEPNSSSTSAFETWIL KLEQVNVYFLDITDLYRDTVARFVLETIARVPVFAFMSVLHMYETFGWR RADYLVKHFASWNEHMLLIMEELGGNSWDFDLAQHATFYFMVLYLILSPRM AVHSECVESHAYTDYKFLKASGEELKNMPADIAVKYITGGDLYLFENLYDFVNI RDEAECHKTHRACTGLSLRSHLEDDDDTEESGCVVPEAHCEGIVDCLKKSIT S"		
intron	complement(6883..7267)		
	/number=1		
exon	complement(7268..7339)		
	/gene="AT4g22260"		
	/number=2		
intron	complement(7340..7506)		
	/number=2		
exon	complement(7507..7575)		
	/gene="AT4g22260"		
	/number=3		
intron	complement(7576..7666)		
	/number=3		
exon	complement(7667..7775)		
	/gene="AT4g22260"		
	/number=4		
intron	complement(7776..7862)		
	/number=4		
exon	complement(7863..7978)		
	/gene="AT4g22260"		
	/number=5		
intron	complement(7979..8360)		
	/number=5		
exon	complement(8361..8454)		
	/gene="AT4g22260"		
	/number=6		
intron	complement(8455..8587)		
	/number=6		
exon	complement(8588..8777)		
	/gene="AT4g22260"		
	/number=7		
intron	complement(8778..8908)		
	/number=7		
exon	complement(8909..9057)		
	/gene="AT4g22260"		
	/number=8		
exon	10103..10244		
	/gene="AT4g22270"		
	/number=1		
gene	10103..12393		
	/gene="AT4g22270"		
CDS	join(10103..10244,10395..10516,11327..11779,11868..12174,12254..12393)		
	/gene="AT4g22270"		
	/note="contains EST gb:AI997575.1"		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="CAB79182.1"		
	/db_xref="GI:7269073"		
	/translation="MATNGSSEHLLPETTPRSALINKQOPEEPASRFTMSLVLFQ SNFGISFVSLSIWSKTKGMRFLDKLMDVSDKRYEYAEIORSKRLMIFVLP SL TEATYRIWYISGPNQIPIIINILSHVVACTIQLSSWLTNRSLFIIVCLYKITCH LQTLRLDDFACFASEITDVSALGEGHQIKRNLRIVSHRFRFILLSLILVATQFM ALLTTTRASAVNIYEGELALCSLSVTGVICLRSAKITHKAQSVTLAAKNVC ATVDSFDLDEETPTGSIIESQVSLRGNAIETSDDEGEGLDNDKIHPIYANTIS YQKRQALVTLENNKAGITVYGLVDRSWLNTFGIELALLLNLNKTIGLA"		
	10245..10394		
intron	/gene="AT4g22270"		
	/number=1		
exon	10395..10516		
	/gene="AT4g22270"		
	/number=2		
intron	10517..11326		
	/gene="AT4g22270"		
	/number=2		
tRNA	11168..11253		
	/gene="AT4g22270"		
	/note="Tyr-tRNA_(GTA)"		
exon	11327..11779		
	/gene="AT4g22270"		
	/number=3		
intron	11780..11867		
	/gene="AT4g22270"		
	/number=3		
exon	11868..12174		
	/gene="AT4g22270"		
	/number=4		
intron	12175..12253		
	/gene="AT4g22270"		

exon /number=4  
12254..12393  
/gene="AT4g22270"  
/number=5  
12761..15111  
/gene="AT4g22280"  
exon complement(12761..13015)  
/gene="AT4g22280"  
/number=1  
complement(join(12761..13015,13081..13227,13379..14157,  
14209..14314,14995..15111))  
/gene="AT4g22280"  
gene complement(join(12761..13015,13081..13227,13379..14157,  
14209..14314,14995..15111))  
/gene="AT4g22280"  
/note="similarity to hypothetical protein, chr.4,  
Arabidopsis thaliana, PATCH:G2392763  
contains EST gb:AI998861.1, AI993240.1"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB79183.1"  
/db\_xref="GI:7269074"  
/translation="MKQVMTDTCIVGRNPATIVNVGLNISHQCMISLKKCVIIF  
EGVVADGDLFLRLITPERFIRIQWRDRINEATSTSVLSHRWYLFAPRNLCLDQE  
VGGDSFIDFVRLVVTGNFPRIIRISIKRMSIDTGVTRWVVDVLEHGVSYLDIDI  
ISDDIGFVPLEIFTCKTLVELKLARGFVAMVPDYVSLPSLKTSLFSSYFNCRDCCP  
LGRLLSACPVELETTIGGCWHIEFCFTVSSSTLKKLTITTSFHYEYMGITDTPSL  
AVLEYSDLPVHPFVNLESVEAKLALDLCSDSNPTNLKGLRNVLEVLSSGYTSQ  
ILYDFEATPVRLSLSKNSRHYIWNLLPILLESKPLETLVIKGPLSADRYE  
REGLSCPVKVLEITYGGKYEELEQMEHFLKLLPCLLELVKVRASALNDEKRSRTND  
LLMVPSSNCKNLKFC"  
intron complement(13016..13080)  
/number=1  
complement(13081..13227)  
exon /gene="AT4g22280"  
/number=2  
complement(13228..13378)  
/number=2  
complement(13379..14157)  
exon /gene="AT4g22280"  
/number=3  
complement(14158..14208)  
intron complement(14209..14314)  
/gene="AT4g22280"  
/number=4  
complement(14315..14994)  
intron complement(14995..15111)  
exon /gene="AT4g22280"  
/number=5  
Query Match 81.7%; Score 18.8; DB 8; Length 199577;  
Best Local Similarity 90.9%; Pred. No. le=02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 aaatccatgtataagcagcagg 23  
|||||  
Db 159343 AAATCCATGTAAACGAGTAAGG 159322  
RESULT 8  
AC072048/c AC072048 208684 bp DNA HTG 07-JUN-2000  
LOCUS Mus musculus chromosome 6 clone RP23-189E15 strain C57BL6/J,  
DEFINITION WORKING DRAFT SEQUENCE, 24 unordered pieces.  
AC072048  
AC072048  
AC072048.1 GI:8313202  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE AUTHORS  
1 (bases 1 to 208684)  
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,  
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,  
Huang,M.C., Idol,J., Lee-Lin,S.-O., Maduro,O.L., Maduro,V.B.,  
Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,  
Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,  
Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,  
Wetherby,K.D. and Green,E.D.  
NISC Mouse Sequencing Initiative  
Unpublished  
2 (bases 1 to 208684)  
Green,E.D.  
Direct Submission  
Submitted (07-JUN-2000) NIH Intramural Sequencing Center, 8717  
Grovemont Circle, Gaithersburg, MD 20877, USA  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.mouse@nih.gov  
----- Project Information  
Center project name: wp  
Center clone name: 189E15  
----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 195465 bases at least Q40  
Consensus quality: 197294 bases at least Q30  
Consensus quality: 198446 bases at least Q20  
Insert size: 209000; agarose-fp  
Insert size: 197000; pulse-field-gel  
Insert size: 206384; sum-of-contigs  
Quality coverage: 6.43x in Q20 bases; agarose-fp  
Quality coverage: 6.82x in Q20 bases; pulse-field-gel  
Quality coverage: 6.51x in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 24 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2209: contig of 2209 bp in length  
\* 2210 2309: gap of unknown length  
\* 2310 4736: contig of 2427 bp in length  
\* 4737 4836: gap of unknown length  
\* 4837 7023: contig of 2187 bp in length  
\* 7024 7123: gap of unknown length  
\* 7124 9505: contig of 2382 bp in length  
\* 9506 9605: gap of unknown length  
\* 9606 14706: contig of 5101 bp in length  
\* 14707 14806: gap of unknown length  
\* 14807 18659: contig of 3853 bp in length  
\* 18660 18759: gap of unknown length  
\* 18760 25424: contig of 6665 bp in length  
\* 25425 25524: gap of unknown length  
\* 25525 33299: contig of 7775 bp in length  
\* 33300 33399: gap of unknown length  
\* 33400 39449: contig of 6050 bp in length  
\* 39450 39550: gap of unknown length  
\* 39551 48515: contig of 8966 bp in length  
\* 48516 48615: gap of unknown length  
\* 48616 55023: contig of 6408 bp in length  
\* 55024 55123: gap of unknown length  
\* 55124 64442: contig of 9319 bp in length  
\* 64443 64542: gap of unknown length  
\* 64543 73027: contig of 8485 bp in length  
\* 73028 73127: gap of unknown length  
\* 73128 82147: contig of 9020 bp in length  
\* 82148 82247: gap of unknown length

\* 82248 93854: contig of 11607 bp in length  
\* 93855 93954: gap of unknown length  
\* 93955 103236: contig of 9282 bp in length  
\* 103236 103237: gap of unknown length  
\* 103237 117502: contig of 14166 bp in length  
\* 117502 117503: gap of unknown length  
\* 117503 130162: contig of 12560 bp in length  
\* 130162 130263: gap of unknown length  
\* 130263 140708: contig of 10446 bp in length  
\* 140708 140809: gap of unknown length  
\* 140809 151652: contig of 10844 bp in length  
\* 151652 151753: gap of unknown length  
\* 151753 163824: contig of 12071 bp in length  
\* 163824 163923: gap of unknown length  
\* 163923 178730: contig of 14807 bp in length  
\* 178730 178831: gap of unknown length  
\* 178831 193909: contig of 15079 bp in length  
\* 193909 194009: gap of unknown length  
\* 194009 208684: contig of 14675 bp in length.  
FEATURES  
source  
1. .208684  
location/Qualifiers  
1. .208684  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/chromosome="6"  
/clone\_lib="RPCI mouse BAC library 23"  
1. .2209  
/note="assembly\_fragment"  
2310. .4736  
/note="assembly\_fragment"  
4837. 7023  
/note="assembly\_fragment"  
7124. .9505  
/note="assembly\_fragment"  
9606. .14706  
/note="assembly\_fragment"  
14807. 18659  
/note="assembly\_fragment"  
18760. .25424  
/note="assembly\_fragment"  
25525. .33299  
/note="assembly\_fragment"  
33400. .39449  
/note="assembly\_fragment"  
39550. 48515  
/note="assembly\_fragment"  
48616. .55023  
/note="assembly\_fragment"  
55124. .64442  
/note="assembly\_fragment"  
64543. 73027  
/note="assembly\_fragment"  
73128. .82147  
/note="assembly\_fragment"  
82248. 93854  
/note="assembly\_fragment"  
93955. 103236  
/note="assembly\_fragment"  
103337. 117502  
/note="assembly\_fragment"  
117603. 130162  
/note="assembly\_fragment"  
130263. 140708  
/note="assembly\_fragment"  
140809. 151652  
/note="assembly\_fragment"  
151753. 163823  
/note="assembly\_fragment"  
163924. 178730  
/note="assembly\_fragment"  
clone\_end.T7  
vector\_side:right"

misc\_feature 178831. .193909  
/note="assembly\_fragment"  
clone\_end.SP6  
vector\_side:right"  
misc\_feature 194010. .208684  
/note="assembly\_fragment"  
BASE COUNT 54390 a 48824 c 48605 g 54542 t 2323 others  
ORIGIN  
Query Match 81.7%; Score 18.8; DB 2; Length 208684;  
Best Local Similarity 90.9%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 aaatccatgtaaagcagcagg 23  
||||| ||||| ||||| |||||  
Db 135605 AAATCCCTGTAAACACAGG 135584  
RESULT 9  
AC046148  
LOCUS  
DEFINITION  
AC046148 168524 bp DNA HTG 06-NOV-2000  
Mus musculus chromosome 14 clone RP23-265D21, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 45 unordered pieces.  
AC046148  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus  
house mouse.  
HTG: HTGS\_PHASE1.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 168524)  
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,  
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,  
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,  
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,  
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,  
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,  
Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,  
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,  
Kovar,C., Liu,J., Liu,W., Louissegh,H., Lozado,R.J., Martin,R.,  
Massey,E., Mcleod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,  
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,  
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,  
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgeson,A.,  
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,  
Worley,K. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 168524)  
Worley,K.C.  
Direct Submission  
Submitted (13-APR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 4, 2000 this sequence version replaced gi:9929612.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: MAEB  
Center clone name: RP23-265D21  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-primer Bodipy; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 132230 bases at least Q40  
Consensus quality: 149333 bases at least Q30  
Consensus quality: 156096 bases at least Q20  
Estimated insert size: 153798; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1  
\* 10853: contig of 10853 bp in length  
\* 10854: gap of unknown length  
\* 10954 20346: contig of 9393 bp in length  
\* 20347 20446: gap of unknown length  
\* 20447 30391: contig of 9945 bp in length  
\* 30392 30491: gap of unknown length  
\* 30492 39349: contig of 8858 bp in length  
\* 39350 39449: gap of unknown length  
\* 39450 47183: contig of 7734 bp in length  
\* 47184 47283: gap of unknown length  
\* 47284 53762: contig of 6479 bp in length  
\* 53763 53862: gap of unknown length  
\* 53863 58669: contig of 4807 bp in length  
\* 58670 58769: gap of unknown length  
\* 58770 65058: contig of 6289 bp in length  
\* 65059 65158: gap of unknown length  
\* 65159 69830: contig of 4672 bp in length  
\* 69831 69930: gap of unknown length  
\* 69931 73910: contig of 3980 bp in length  
\* 73911 74010: gap of unknown length  
\* 74011 78684: contig of 4674 bp in length  
\* 78685 78784: gap of unknown length  
\* 78785 83343: contig of 4559 bp in length  
\* 83344 83443: gap of unknown length  
\* 83444 87849: contig of 4406 bp in length  
\* 87850 87949: gap of unknown length  
\* 87950 92076: contig of 4127 bp in length  
\* 92077 92176: gap of unknown length  
\* 92177 97227: contig of 5051 bp in length  
\* 97228 97327: gap of unknown length  
\* 97328 101356: contig of 4029 bp in length  
\* 101357 101456: gap of unknown length  
\* 101457 104693: contig of 3237 bp in length  
\* 104694 104793: gap of unknown length  
\* 104794 108132: contig of 3339 bp in length  
\* 108133 108233: gap of unknown length  
\* 108233 111747: contig of 3515 bp in length  
\* 111748 111847: gap of unknown length  
\* 111848 115180: contig of 3333 bp in length  
\* 115181 115280: gap of unknown length  
\* 115281 118248: contig of 3288 bp in length  
\* 118249 118549: gap of unknown length  
\* 118549 121539: contig of 2891 bp in length  
\* 121540 121639: gap of unknown length  
\* 121640 124526: contig of 2887 bp in length  
\* 124527 124627: gap of unknown length  
\* 124627 127513: contig of 2886 bp in length  
\* 127513 127612: gap of unknown length  
\* 127613 131027: contig of 3415 bp in length  
\* 131028 131127: gap of unknown length  
\* 131128 131604: contig of 2477 bp in length  
\* 131605 133704: gap of unknown length  
\* 133705 135994: contig of 2190 bp in length  
\* 135994 137407: gap of unknown length  
\* 137408 137507: gap of unknown length  
\* 137508 139522: contig of 2015 bp in length  
\* 139523 139622: gap of unknown length  
\* 139623 141182: contig of 1560 bp in length  
\* 141183 141282: gap of unknown length  
\* 141283 143104: contig of 1822 bp in length

\* 143105 143204: gap of unknown length  
\* 143205 145597: contig of 2393 bp in length  
\* 145598 145697: gap of unknown length  
\* 145698 147840: contig of 2143 bp in length  
\* 147841 147940: gap of unknown length  
\* 147941 150388: contig of 2448 bp in length  
\* 150389 150488: gap of unknown length  
\* 150489 151853: contig of 1365 bp in length  
\* 151854 151953: gap of unknown length  
\* 151954 154894: contig of 2941 bp in length  
\* 154895 154994: gap of unknown length  
\* 154995 156000: contig of 1606 bp in length  
\* 156001 156700: gap of unknown length  
\* 156701 158135: contig of 1435 bp in length  
\* 158136 158235: gap of unknown length  
\* 158236 159267: contig of 1032 bp in length  
\* 159268 159367: gap of unknown length  
\* 159368 161428: contig of 2061 bp in length  
\* 161429 161528: gap of unknown length  
\* 161529 162993: contig of 1465 bp in length  
\* 162994 163093: gap of unknown length  
\* 163094 164386: contig of 1293 bp in length  
\* 164387 164486: gap of unknown length  
\* 164487 165850: contig of 1364 bp in length  
\* 165851 165950: gap of unknown length  
\* 165951 167364: contig of 1414 bp in length  
\* 167365 167464: gap of unknown length  
\* 167465 168524: contig of 1060 bp in length.

#### FEATURES

Location/Qualifiers  
1. .168524  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="14"

BASE COUNT 45461 a 36455 c 36516 g 45685 t 4407 others  
ORIGIN

Query Match 80.0%; Score 18.4; DB 2; Length 168524;  
Best Local Similarity 95.0%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aaatccatgtataagcagcag 21  
|||||

Db 112558 AAATCCATGTATACACGAG 112577

RESULT 10  
AC040920/c

LOCUS

AC040920 172650 bp DNA HTG 11-APR-2000  
Homo sapiens chromosome 5 clone RP11-509M23 map 5, WORKING DRAFT  
SEQUENCE, 19 unordered pieces.

DEFINITION

AC040920.1 GI:7534094

VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS

human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 172650)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 5, clone RP11-509M23

Unpublished

2 (bases 1 to 172650)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Lieu, C., Liu, G., Locke, C., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrim, J., Menus, L., Minova, T., Miranda, C., Mienga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody M.  
 Direct Submission

# TITLE JOURNAL

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8553

Center clone name: 509\_M\_23

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 163069 bases at least Q40

Consensus quality: 167976 bases at least Q30

Consensus quality: 169595 bases at least Q20

Insert size: 170000; agarose-fp

Quality coverage: 4.7 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
 consists of 19 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 422: contig of 422 bp in length  
 423 522: gap of 100 bp  
 523 2269: contig of 1747 bp in length  
 2270 2369: gap of 100 bp  
 2370 4946: contig of 2577 bp in length  
 4947 5046: gap of 100 bp  
 5047 7162: contig of 2116 bp in length  
 7163 7262: gap of 100 bp  
 7263 9787: contig of 2525 bp in length  
 9788 9887: gap of 100 bp  
 9888 13694: contig of 3807 bp in length  
 13695 13794: gap of 100 bp  
 13795 18462: contig of 4688 bp in length  
 18483 18582: gap of 100 bp  
 18583 24380: contig of 5798 bp in length  
 24381 24480: gap of 100 bp  
 24481 30597: contig of 6117 bp in length  
 30598 30697: gap of 100 bp  
 30698 36241: contig of 5544 bp in length  
 36242 36341: gap of 100 bp  
 36342 43886: contig of 7545 bp in length  
 43887 43986: gap of 100 bp  
 43987 53642: contig of 9656 bp in length  
 53643 53742: gap of 100 bp  
 53743 65843: contig of 12101 bp in length  
 65844 65943: gap of 100 bp  
 65944 82865: contig of 16922 bp in length

\* 82866 82965: gap of 100 bp  
 82966 97230: contig of 14265 bp in length  
 97231 97330: gap of 100 bp  
 97331 114682: contig of 17352 bp in length  
 114683 114782: gap of 100 bp  
 114783 132889: contig of 17507 bp in length  
 132890 132389: gap of 100 bp  
 132390 151664: contig of 19275 bp in length  
 151665 151764: gap of 100 bp  
 151765 172650: contig of 20886 bp in length.  
 Location/Qualifiers  
 1. .172650  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /map="5"  
 /clone="RP11-509M23"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. .422  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:left  
 523. .2269  
 /note="assembly\_fragment"  
 2370. .4946  
 /note="assembly\_fragment"  
 5047. .7162  
 /note="assembly\_fragment"  
 7263. .9787  
 /note="assembly\_fragment"  
 9888. .13694  
 /note="assembly\_fragment"  
 13795. .18482  
 /note="assembly\_fragment"  
 18583. .24380  
 /note="assembly\_fragment"  
 24481. .30597  
 /note="assembly\_fragment"  
 30698. .36241  
 /note="assembly\_fragment"  
 36342. .43886  
 /note="assembly\_fragment"  
 43987. .53642  
 /note="assembly\_fragment"  
 53743. .65843  
 /note="assembly\_fragment"  
 65944. .82865  
 /note="assembly\_fragment"  
 82966. .97230  
 /note="assembly\_fragment"  
 97331. .114682  
 /note="assembly\_fragment"  
 114783. .132889  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 132390. .151664  
 /note="assembly\_fragment"  
 151765. .172650  
 /note="assembly\_fragment"  
 1804 others

## FEATURES source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

Query Match 80.0%; Score 18.4; DB 2; Length 172650;  
 Best Local Similarity 95.0%; Pred. No. 1.6e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aaatccatgtaaacgacgac 21

|||||

Db 142729 AAAGCATGTAAAGCAGCAG 142710

```

RESULT 11
AF118429          535 bp      DNA          VRT          10-JAN-2001
LOCUS             Falc0 peregrinus microsatellite NVH Fp86-2 sequence.
DEFINITION
ACCESSION         AF118429
VERSION           AF118429.1  GI:5840932
KEYWORDS
SOURCE
ORGANISM          Falc0 peregrinus.
                  Falc0 peregrinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Falconiformes; Falconidae; Falco.
1 (bases 1 to 535)
Nesje,M., Roed,K.H., Lifjeld,J.T., Lindberg,P. and Steen,O.F.
Genetic relationships in the peregrine falcon (Falc0 peregrinus)
analysed by microsatellite DNA markers
Mol. Ecol. 9 (1), 53-60 (2000)
JOURNAL           20117586
MEDLINE           10652075
REFERENCE
AUTHORS           Nesje,M. and Roed,K.H.
TITLE             Direct Submission
JOURNAL           Submitted (07-JAN-1999) MGA, Genetics, Norwegian College of
Veterinary Medicine, P.O. Box 8146 Dep., Oslo 0033, Norway
FEATURES
source            1..535
                  /organism="Falc0 peregrinus"
                  /db_xref="taxon:8954"
repeat_region    1..535
                  /note="microsatellite NVH Fp86-2"
                  /rpt_type=tandem
BASE COUNT        168 a 110 c 117 g 140 t
ORIGIN
Query Match      79.1%; Score 18.2; DB 5; Length 535;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gaaatccatgtaaagcagcaggg 23
| ||||| ||||| ||||| |||||
Db 385 GGAATCCATGACACGACGAGG 407

RESULT 12
AC005083/c
LOCUS             AC005083 146285 bp      DNA          PRI          21-DEC-1999
DEFINITION        Homo sapiens BAC clone CTA-281G5 from 7p15-p21, complete sequence.
ACCESSION         AC005083
VERSION           AC005083.1  GI:4150930
KEYWORDS
SOURCE            human.
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 146285)
Madsen,C. and Blair,T.
The sequence of Homo sapiens BAC clone CTA-281G5
JOURNAL           Unpublished
REFERENCE
AUTHORS           2 (bases 1 to 146285)
                  Waterston,R.H.
TITLE             Direct Submission
JOURNAL           Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 146285)
Waterston,R.
Direct Submission
JOURNAL           Submitted (12-JAN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 146285)
Waterston,R.
Direct Submission
JOURNAL           Submitted (21-DEC-1999) Department of Genetics, Washington

```

## COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Jan 12, 1999 this sequence version replaced gi:3212908.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_RG281G05  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send [mailto:egreen@nhgri.nih.gov](mailto:mailto:egreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

Clone CTA-281G5 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBel0BAC11  
Selection: chloramphenicol

## NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-281G5; actual end is at 146285 of CTA-281G5

BAC CTA-281G5 contains an E. coli transposon from 10982 to 12345 that is not represented in the submitted sequence.

## FEATURES

## source

```

1..146285
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p15-p21"
/clone="CTA-281G5"
/clone_lib="CITB-HS-A"
322..358
/rpt_family="(TAAA)n"
808..952
/rpt_family="MER1_type"
1172..1194
/rpt_family="AT-rich"
1264..1720
/rpt_family="L1"
1848..1881
/rpt_family="AT-rich"
1882..2171
/rpt_family="Alu"
3782..3804
/rpt_family="AT-rich"
4794..4826
/rpt_family="AT-rich"
5764..6034

```

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

## repeat\_region

```

repeat_region      /rpt_family="purine-rich"
6891..7126
repeat_region      /rpt_family="L1"
7133..7263
repeat_region      /rpt_family="L1"
7455..7481
repeat_region      /rpt_family="AT-rich"
7504..7554
repeat_region      /rpt_family="AT-rich"
8320..8348
repeat_region      /rpt_family="(CA)n"
8466..8533
repeat_region      /rpt_family="AT-rich"
9611..9733
repeat_region      /rpt_family="MIR"
9946..10172
repeat_region      /rpt_family="MER1_type"
10437..10477
repeat_region      /rpt_family="(CA)n"
10581..10681
repeat_region      /rpt_family="(TAAA)n"
10759..11049
repeat_region      /rpt_family="AT-rich"
11050..11235
repeat_region      /rpt_family="(CAATA)n"
11408..11526
repeat_region      /rpt_family="(CATAT)n"
11768..11872
repeat_region      /rpt_family="(CATTA)n"
12002..12110
repeat_region      /rpt_family="(CATAT)n"
12057..12168
repeat_region      /rpt_family="(TA)n"
12381..12490
repeat_region      /rpt_family="(CAATA)n"
12562..12680
repeat_region      /rpt_family="(CTATA)n"
13131..13367
repeat_region      /rpt_family="L2"
13378..13540
repeat_region      /rpt_family="MER1_type"
13557..13744
repeat_region      /rpt_family="L2"
13745..14592
repeat_region      /rpt_family="L1"
14593..14616
repeat_region      /rpt_family="L2"
16078..16548
repeat_region      /rpt_family="L1"
18492..18692
repeat_region      /rpt_family="MER1_type"
19334..19422
repeat_region      /rpt_family="LFR41"
19880..20137
repeat_region      /rpt_family="LFR41"
20504..20541
repeat_region      /rpt_family="MER1_type"
21092..21607
repeat_region      /rpt_family="MaLR"
22221..22499
repeat_region      /rpt_family="Alu"
22500..22523
repeat_region      /rpt_family="(CAAAA)n"
22535..22857
repeat_region      /rpt_family="MER4-group?"
23203..23380
repeat_region      /rpt_family="L1"
23381..23643
repeat_region      /rpt_family="Alu"
23644..24770
repeat_region      /rpt_family="L1"
25826..25846
repeat_region      /rpt_family="AT-rich"

```

```

repeat_region      26130..26254
/rpt_family="MIR"
repeat_region      26529..26563
/rpt_family="AT-rich"
repeat_region      26792..27116
/rpt_family="Alu"
repeat_region      30744..30982
/rpt_family="MIR"
repeat_region      31032..31739
/rpt_family="L1"
repeat_region      32332..32448
/rpt_family="MIR"
repeat_region      32536..32730
/rpt_family="MIR"
repeat_region      33282..33333
/rpt_family="AT-rich"
repeat_region      35261..35288
/rpt_family="(CATTT)n"
repeat_region      35571..35598
/rpt_family="(CA)n"
repeat_region      35924..36007
/rpt_family="L2"
repeat_region      36476..37873
/rpt_family="L1"
repeat_region      38324..38899
/rpt_family="Retroviral"

```

```

Query Match      79.1%; Score 18.2; DB 9; Length 146285;
Best Local Similarity 87.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaagcagcagg 23
|| ||||| ||||| |||||
Db 20614 GACATCCATGTAAGGAGCAGG 20592

RESULT 13
AC023167/c
LOCUS      AC023167 183634 bp DNA
DEFINITION Mus musculus clone RP23-361K18, *** SEQUENCING IN PROGRESS ***, 61
unorderded pieces.
ACCESSION  AC023167
VERSION     AC023167.8 GI:11079356
KEYWORDS   HTG: HTGS-PHASE1.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hoques, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loutsched, H., Lozardo, R. J., Martin, R.,
Massey, E., McLeod, M. P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Oguh, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,
Worley, K. and Gibbs, R.
Direct Submission
TITLE
JOURNAL
REFERENCE  2 (bases 1 to 183634)
AUTHORS   Worley, K.C.
Direct Submission
Submitted (09-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Nov 3, 2000 this sequence version replaced gi:8248589.

```



```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAAW
Center clone name: RP23-361K18
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-primer Bodipy: 70% of reads
Chemistry: Dye-terminator Big Dye: 30% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 137454 bases at least Q40
Consensus quality: 161084 bases at least Q30
Consensus quality: 170194 bases at least Q20
Estimated insert size: 164611; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 10583: contig of 10583 bp in length
* 10584: gap of unknown length
* 10684: contig of 9744 bp in length
* 20428: gap of unknown length
* 20528: contig of 8536 bp in length
* 29064: gap of unknown length
* 29164: contig of 7048 bp in length
* 36212: gap of unknown length
* 36312: contig of 5943 bp in length
* 42255: gap of unknown length
* 42355: contig of 5338 bp in length
* 47693: gap of unknown length
* 51859: contig of 4067 bp in length
* 51959: gap of unknown length
* 51960: contig of 3041 bp in length
* 55001: gap of unknown length
* 55101: contig of 4666 bp in length
* 59767: gap of unknown length
* 59867: contig of 2791 bp in length
* 62658: gap of unknown length
* 62758: contig of 4413 bp in length
* 67171: gap of unknown length
* 67270: contig of 2929 bp in length
* 70199: gap of unknown length
* 70299: gap of unknown length
* 70300: contig of 3826 bp in length
* 74126: gap of unknown length
* 74226: contig of 3082 bp in length
* 77307: gap of unknown length
* 77407: contig of 2196 bp in length
* 79603: gap of unknown length
* 79704: contig of 2982 bp in length
* 82685: gap of unknown length
* 82786: contig of 3599 bp in length
* 86385: gap of unknown length
* 86485: contig of 3583 bp in length
* 90067: gap of unknown length
* 90167: contig of 2851 bp in length
* 93018: gap of unknown length
* 93118: contig of 3875 bp in length
* 96993: gap of unknown length
* 97093: contig of 2919 bp in length
* 100012: gap of unknown length
* 100112: contig of 3745 bp in length
* 100113: gap of unknown length
*
* 103858: gap of unknown length
* 103958: contig of 3565 bp in length
* 107522: gap of unknown length
* 107623: contig of 2424 bp in length
* 110046: gap of unknown length
* 110147: contig of 2952 bp in length
* 113098: gap of unknown length
* 113198: contig of 3372 bp in length
* 113199: gap of unknown length
* 116571: contig of 2343 bp in length
* 116670: gap of unknown length
* 119013: contig of 2940 bp in length
* 119114: gap of unknown length
* 122053: contig of 2806 bp in length
* 122154: gap of unknown length
* 124960: contig of 2109 bp in length
* 125060: gap of unknown length
* 127168: contig of 2336 bp in length
* 127268: gap of unknown length
* 127269: contig of 3451 bp in length
* 129605: gap of unknown length
* 129705: contig of 3353 bp in length
* 133156: gap of unknown length
* 133255: contig of 1419 bp in length
* 133256: gap of unknown length
* 136609: contig of 2776 bp in length
* 136709: gap of unknown length
* 138128: contig of 2591 bp in length
* 138228: gap of unknown length
* 141004: contig of 1117 bp in length
* 141104: gap of unknown length
* 143695: contig of 1349 bp in length
* 143795: gap of unknown length
* 144912: contig of 2530 bp in length
* 145012: gap of unknown length
* 146361: contig of 1458 bp in length
* 146460: gap of unknown length
* 146461: contig of 2092 bp in length
* 148991: gap of unknown length
* 149090: contig of 1921 bp in length
* 149091: gap of unknown length
* 150392: contig of 1710 bp in length
* 150492: gap of unknown length
* 152202: contig of 2092 bp in length
* 152302: gap of unknown length
* 152303: contig of 1471 bp in length
* 154395: gap of unknown length
* 154494: contig of 1467 bp in length
* 154915: gap of unknown length
* 156115: contig of 1458 bp in length
* 156516: gap of unknown length
* 157973: contig of 1776 bp in length
* 158073: gap of unknown length
* 159544: contig of 1471 bp in length
* 159644: gap of unknown length
* 161111: contig of 1612 bp in length
* 161212: gap of unknown length
* 162823: contig of 1641 bp in length
* 162923: gap of unknown length
* 164564: contig of 1431 bp in length
* 164664: gap of unknown length
* 166095: contig of 1431 bp in length
* 166195: gap of unknown length
* 166196: contig of 1776 bp in length
* 167971: gap of unknown length
* 168071: contig of 1235 bp in length
* 169306: gap of unknown length
* 169406: contig of 1288 bp in length
* 170694: gap of unknown length
* 170794: contig of 1559 bp in length
* 172353: gap of unknown length
* 172453: contig of 1010 bp in length
* 173463: gap of unknown length
* 173563: contig of 1339 bp in length
* 174902: gap of unknown length
* 174903: contig of 1924 bp in length
* 175002: gap of unknown length
* 175003: contig of 1483 bp in length
* 176926: gap of unknown length
* 177026: contig of 1225 bp in length
* 177027: gap of unknown length
* 178510: contig of 1225 bp in length
* 178609: gap of unknown length
* 179834: contig of 1225 bp in length
* 179835: gap of unknown length

```

\* 179935 181235: contig of 1301 bp in length  
\* 181236 181335: gap of unknown length  
\* 181336 182487: contig of 1152 bp in length  
\* 182488 182587: gap of unknown length  
\* 182588 183634: contig of 1047 bp in length.

FEATURES  
source  
1. 183634  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-361K18"  
BASE COUNT 48869 a 38982 c 39763 g 49962 t 6058 others  
ORIGIN

Query Match 79.1%; Score 18.2; DB 2; Length 183634;  
Best Local Similarity 87.0%; Pred No. 2.1e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 gaatccatgttaagcagcagg 23  
||||| ||| ||||| |||||  
Db 60779 GAAATCCCTGCTAGCAGCAGG 60757

RESULT 14  
AC021443/c  
LOCUS  
DEFINITION Homo sapiens chromosome 11 clone RP11-692M12 map 11, WORKING DRAFT  
AC021443  
AC021443 8 GI:14269788  
VERSION  
KEYWORDS HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 228809)  
Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,  
Anderson S., Baldwin J., Barna N., Beckerly R., Beda F.,  
Boguslavsky L., Bouckgalter B., Brown A., Burkett G., Castle A.,  
DeArelano K., Dewar K., Domino M., Doyle M., Penestor J.,  
Pereira P., Fitzhugh W., Forrest C., Gage D., Galagan J.,  
Gardyna S., Grant G., Hagos B., Heaford A., Horton L.,  
Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,  
Lander T., Lehoczy J., Levine R., Lieu C., Liu G., Locke K.,  
Macdonald P., Marquis J., McEwan P., McGurk A., McKernan K.,  
McPheeters R., Meldrim J., Meneus L., Morrow J., Naylor J.,  
Norman C.H., O'Connor T., O'Donnell P., Olivari T.M., Peterson K.,  
Pierre N., Pisani C., Pollara V., Raymond C., Riley R., Rothman D.,  
Roy A., Santos R., Severi P., Spencer B., Stange-Thomann N.,  
Stojanovic N., Subramanian A., Talamas J., Testaye S., Theodore J.,  
Tirrell A., Vassiliou H., Viel R., Vo A., Wu X., Wyman D., Ye W.J.,  
Zimmer A. and Zody M.

Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research 320 Charles Street, Cambridge, MA 02141, USA  
On May 31, 2001 this sequence version replaced gi:11560245.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: W1BR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L5123  
Center clone name: 692\_M12  
----- Summary Statistics

\*\*\*\*\*  
TITLE  
JOURNAL  
COMMENT

Sequencing vector: M13; M7815; 43% of reads  
Sequencing vector: Plasmid; L08752; 1% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 214328 bases at least Q40  
Consensus quality: 219470 bases at least Q30  
Consensus quality: 222925 bases at least Q20  
Insert size: 198000; agarose-fp  
Insert size: 226409; sum-of-contigs  
Quality coverage: 10.4 in Q20.  
NOTE: This is a 'working draft' sequence. It currently  
consists of 25 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 3785: contig of 3785 bp in length  
3786 3885: gap of 100 bp  
3886 5107: contig of 1222 bp in length  
5108 5207: gap of 100 bp  
5208 6375: contig of 1168 bp in length  
6376 6475: gap of 100 bp  
6476 7525: contig of 1050 bp in length  
7526 7625: gap of 100 bp  
7626 8701: contig of 1076 bp in length  
8702 8801: gap of 100 bp  
8802 10296: contig of 1495 bp in length  
10297 10396: gap of 100 bp  
10397 11470: contig of 1074 bp in length  
11471 11570: gap of 100 bp  
11571 12684: contig of 1114 bp in length  
12685 12784: gap of 100 bp  
12785 14999: contig of 2215 bp in length  
15000 15099: gap of 100 bp  
15100 18672: contig of 3573 bp in length  
18673 18772: gap of 100 bp  
18773 22303: contig of 3531 bp in length  
22304 22403: gap of 100 bp  
22404 28525: contig of 6122 bp in length  
28526 28625: gap of 100 bp  
28626 36093: contig of 7468 bp in length  
36094 36193: gap of 100 bp  
36194 45198: contig of 9005 bp in length  
45199 45298: gap of 100 bp  
45299 56916: contig of 11618 bp in length  
56917 57016: gap of 100 bp  
57017 65985: contig of 8969 bp in length  
65986 66085: gap of 100 bp  
66086 78887: contig of 12802 bp in length  
78888 78987: gap of 100 bp  
78988 94277: contig of 15290 bp in length  
94278 94377: gap of 100 bp  
94378 109994: contig of 15617 bp in length  
109995 110094: gap of 100 bp  
110095 121149: contig of 11055 bp in length  
121150 121249: gap of 100 bp  
121250 138125: contig of 16876 bp in length  
138126 138225: gap of 100 bp  
138226 155392: contig of 17167 bp in length  
155393 155492: gap of 100 bp  
155493 179217: contig of 23725 bp in length  
179218 179317: gap of 100 bp  
179318 202705: contig of 23388 bp in length  
202706 202805: gap of 100 bp  
202806 228809: contig of 26004 bp in length.

Location/Qualifiers  
1. 228809  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="11"

FEATURES  
source

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:38 ; Search time 716.55 Seconds  
(without alignments)  
27.519 Million cell updates/sec

Title: US-09-698-903B-11

Perfect score: 23

Sequence: 1 gaaatccatgtaaacgacgagg 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	AAD07000	PCR primer B01, to
2	23	100.0	415	AAD06997	Right (5') border
3	23	100.0	416	AAD06999	Left (3') border f
4	18.8	81.7	1152	AAC46330	Arabidopsis thalia
5	17.4	75.7	1567	AAAC79889	Human secreted pro
6	17.2	74.8	1153	AAC39525	Arabidopsis thalia
7	17.2	74.8	2857	AAH18005	Human cDNA sequenc
8	17.2	74.8	4495	AAI58467	Human polynucleoti
9	17.2	74.8	4775	AAI60253	Human polynucleoti
10	17.2	74.8	8826	AAAT74989	Saccharomyces cere
11	17.2	74.8	9621	AAT94548	Mutant YLR087c gen

c 12	16.8	73.0	602	22	AAH29132	Drosophila melanog
c 13	16.8	73.0	1752	22	AAF57408	Human p95 protein
c 14	16.8	73.0	4388	18	AAT60073	Selective marker g
15	16.6	72.2	281	14	AAQ60834	Human brain Expres
c 16	16.6	72.2	292	16	AAT25120	Human gene signatu
c 17	16.6	72.2	292	20	AAV88563	EST clone EY281.
c 18	16.6	72.2	292	21	AAZ46937	Human saccharide-t
c 19	16.6	72.2	549	19	AAV49585	Human stomach canc
c 20	16.6	72.2	552	21	AAV39940	Human TANGO 183 co
c 21	16.6	72.2	575	22	AAF93941	Primer specific fo
c 22	16.6	72.2	692	22	AAH32272	Human olfactory re
c 23	16.6	72.2	917	21	AAC76546	Human ORFX ORF2101
c 24	16.6	72.2	1026	21	AAC51028	Arabidopsis thalia
c 25	16.6	72.2	1056	21	AAF51420	Arabidopsis thalia
c 26	16.6	72.2	1499	22	AAF93749	Human cDNA encodin
c 27	16.6	72.2	1499	22	AAF29357	Anyloid-beta prote
c 28	16.6	72.2	1504	19	AAV49584	Human stomach canc
c 29	16.6	72.2	1563	22	AAH64828	Human secreted pro
c 30	16.6	72.2	1564	21	AAV37041	Human PRO1248 (UNQ
c 31	16.6	72.2	1564	22	AAF54249	DNA encoding prote
c 32	16.6	72.2	1565	21	AAV39939	Human TANGO 183 cD
c 33	16.6	72.2	1571	20	AAZ41983	Human endometrium
c 34	16.6	72.2	1645	21	AAF21831	Human breast and o
c 35	16.6	72.2	1821	21	AAC51418	Arabidopsis thalia
c 36	16.6	72.2	1851	21	AAC51029	Arabidopsis thalia
c 37	16.6	72.2	1929	22	AAC85075	Atherosclerosis-as
c 38	16.6	72.2	1962	22	AAF33128	Human secreted pro
c 39	16.6	72.2	2370	22	AAF26956	Human cancer assoc
c 40	16.6	72.2	2380	21	AAF76858	Human ORFX ORF2413
c 41	16.6	72.2	2383	22	AAH22608	Human cDNA encodin
c 42	16.6	72.2	2393	22	AAH98580	Human EST-derived
c 43	16.6	72.2	2393	22	AAH22844	Human cDNA encodin
c 44	16.6	72.2	2402	21	AAZ33327	Human secreted pro
c 45	16.6	72.2	2460	20	AAH58671	Human organic cati

#### ALIGNMENTS

##### RESULT 1

AAD07000

ID AAD07000 standard; DNA; 23 BP.

XX

AC AAD07000;

XX

DT 06-AUG-2001 (first entry)

XX

DE PCR primer B01, to recognise foreign DNA and flanking sequence of MS-B2.

XX

KW MS-B2 elite event; transgenic Brassica plant; transformation event;

KW male-sterility gene; PCR primer; ss.

XX

OS Unidentified.

XX

PN WO200131042-A2.

XX

PD 03-MAY-2001.

XX

PF 26-OCT-2000; 2000WO-EP10680.

XX

PR 29-OCT-1999; 99US-0430497.

XX

PA (AVET ) AVENTIS CROPS SCIENCE NV.

XX

PI Weston B, De Beuckeleer M;

XX

DR WPI; 2001-300517/31.

XX

PT Transgenic Brassica plants, seeds, cells or tissues, characterized by harboring specific transformation events, particularly by presence of male-sterility gene, at specific location in its genome -

XX

PS Claim 1; Page 33; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.  
 XX Sequence 23 BP; 9 A; 4 C; 7 G; 3 T; 0 other;  
 SQ

Query Match 100.0%; Score 23; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.088;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaaatccatgtaaagcagcagg 23  
 Db 1 gaaatccatgtaaagcagcagg 23  
 |||||

RESULT 2  
 AAD06997  
 ID AAD06997 standard; DNA; 415 BP.  
 XX AC AAD06997;  
 XX DT 06-AUG-2001 (first entry)  
 XX DE Right (5') border flanking region of elite event MS-B2.  
 XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 XX KW male-sterility gene; ds.  
 XX OS Chimeric - Agrobacterium sp.  
 XX OS Chimeric - Brassica sp.  
 XX FH Key Location/Qualifiers  
 FT misc\_feature 1..234  
 FT /\*tag= a  
 FT /note= "Corresponds to plant DNA"  
 FT misc\_feature 235..415  
 FT /\*tag= b  
 FT /note= "Corresponds to T-DNA"  
 XX WO200131042-A2.  
 XX 03-MAY-2001.  
 XX 26-OCT-2000; 2000WO-EP10680.  
 XX 29-OCT-1999; 99US-0430497.  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX Weston B, De Beuckeleer M;  
 XX WPI; 2001-300517/31.  
 XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX Claim 11; Page 51; 53pp; English.  
 XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the

CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is right (5') border flanking region of elite event  
 CC MS-B2.  
 XX Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;  
 SQ

Query Match 100.0%; Score 23; DB 22; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gaaatccatgtaaagcagcagg 23  
 Db 152 gaaatccatgtaaagcagcagg 174  
 |||||

RESULT 3  
 AAD06999/c  
 ID AAD06999 standard; DNA; 416 BP.  
 XX AC AAD06999;  
 XX DT 06-AUG-2001 (first entry)  
 XX DE Left (3') border flanking region of elite event MS-B2.  
 XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 XX KW male-sterility gene; ds.  
 XX OS Chimeric - Agrobacterium sp.  
 XX OS Chimeric - Brassica sp.  
 XX FH Key Location/Qualifiers  
 FT misc\_feature 1..193  
 FT /\*tag= a  
 FT /note= "Corresponds to T-DNA"  
 FT misc\_feature 194..416  
 FT /\*tag= b  
 FT /note= "Corresponds to plant DNA"  
 XX WO200131042-A2.  
 XX 03-MAY-2001.  
 XX 26-OCT-2000; 2000WO-EP10680.  
 XX 29-OCT-1999; 99US-0430497.  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX Weston B, De Beuckeleer M;  
 XX WPI; 2001-300517/31.  
 XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX Claim 11; Page 52; 53pp; English.  
 XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is left (3') border flanking region of elite event  
 CC MS-B2.  
 XX Sequence 416 BP; 137 A; 72 C; 54 G; 152 T; 1 other;  
 SQ

Query Match	100.0%;	Score 23;	DB 22;	Length 416;
Best Local Similarity	100.0%;	Pred. No. 0.14;		
Matches 23;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	gaatcccatgtaaagcacgagg	23	
Db	263	GAATCCCATGTAAAGCACGAGG	241	
RESULT	4			
AAC46330				
ID	AAC46330	standard; DNA; 1152 BP.		
XX	AC	AAC46330;		
XX	AC			
XX	DT	18-OCT-2000 (first entry)		
XX	XX			
XX	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 49743.		
XX	KW	Hybridisation assay; genetic mapping; gene expression control;		
KW	KW	protein identification; signal transduction pathway;		
KW	KW	metabolic pathway; promoter; termination sequence; ss.		
XX	XX			
OS	Arabidopsis thaliana.			
XX	XX			
PN	EP1033405-A2.			
XX	XX			
XX	PD	06-SEP-2000.		
XX	XX			
PF	25-FEB-2000;	2000EP-0301439.		
XX	XX			
XX	25-FEB-1999;	99US-0121825.		
PR	05-MAR-1999;	99US-0123180.		
PR	09-MAR-1999;	99US-0123548.		
PR	23-MAR-1999;	99US-0125788.		
PR	25-MAR-1999;	99US-0126284.		
PR	29-MAR-1999;	99US-0126785.		
PR	01-APR-1999;	99US-0127462.		
PR	06-APR-1999;	99US-0128234.		
PR	08-APR-1999;	99US-0128714.		
PR	16-APR-1999;	99US-0129845.		
PR	19-APR-1999;	99US-0130077.		
PR	21-APR-1999;	99US-0130449.		
PR	23-APR-1999;	99US-0130510.		
PR	28-APR-1999;	99US-0130891.		
PR	28-APR-1999;	99US-0131449.		
PR	30-APR-1999;	99US-0132048.		
PR	30-APR-1999;	99US-0132407.		
PR	04-MAY-1999;	99US-0132484.		
PR	05-MAY-1999;	99US-0132485.		
PR	06-MAY-1999;	99US-0132486.		
PR	06-MAY-1999;	99US-0132487.		
PR	07-MAY-1999;	99US-0132863.		
PR	11-MAY-1999;	99US-0134236.		
PR	14-MAY-1999;	99US-0134218.		
PR	14-MAY-1999;	99US-0134219.		
PR	14-MAY-1999;	99US-0134221.		
PR	14-MAY-1999;	99US-0134370.		
PR	18-MAY-1999;	99US-0134768.		
PR	19-MAY-1999;	99US-0134941.		
PR	20-MAY-1999;	99US-0135124.		
PR	21-MAY-1999;	99US-0135353.		
PR	24-MAY-1999;	99US-0135629.		
PR	25-MAY-1999;	99US-0136021.		
PR	27-MAY-1999;	99US-0136392.		
PR	28-MAY-1999;	99US-0136782.		
PR	01-JUN-1999;	99US-0137222.		
PR	03-JUN-1999;	99US-0137528.		
PR	04-JUN-1999;	99US-0137502.		
PR	07-JUN-1999;	99US-0137724.		
PR	08-JUN-1999;	99US-0138094.		
PR	10-JUN-1999;	99US-0138540.		
PR	10-JUN-1999;	99US-0138847.		
PR	14-JUN-1999;	99US-0139119.		
PR	16-JUN-1999;	99US-0139452.		
PR	16-JUN-1999;	99US-0139453.		
PR	17-JUN-1999;	99US-0139492.		
PR	18-JUN-1999;	99US-0139454.		
PR	18-JUN-1999;	99US-0139455.		
PR	18-JUN-1999;	99US-0139456.		
PR	18-JUN-1999;	99US-0139457.		
PR	18-JUN-1999;	99US-0139458.		
PR	18-JUN-1999;	99US-0139459.		
PR	18-JUN-1999;	99US-0139460.		
PR	18-JUN-1999;	99US-0139461.		
PR	18-JUN-1999;	99US-0139462.		
PR	18-JUN-1999;	99US-0139463.		
PR	18-JUN-1999;	99US-0139750.		
PR	18-JUN-1999;	99US-0139763.		
PR	21-JUN-1999;	99US-0139817.		
PR	22-JUN-1999;	99US-0139899.		
PR	23-JUN-1999;	99US-0140353.		
PR	23-JUN-1999;	99US-0140354.		
PR	24-JUN-1999;	99US-0140695.		
PR	28-JUN-1999;	99US-0140823.		
PR	29-JUN-1999;	99US-0140991.		
PR	30-JUN-1999;	99US-0141287.		
PR	01-JUL-1999;	99US-0141842.		
PR	01-JUL-1999;	99US-0142154.		
PR	02-JUL-1999;	99US-0142055.		
PR	06-JUL-1999;	99US-0142390.		
PR	08-JUL-1999;	99US-0142803.		
PR	09-JUL-1999;	99US-0142920.		
PR	12-JUL-1999;	99US-0142977.		
PR	13-JUL-1999;	99US-0143542.		
PR	14-JUL-1999;	99US-0143624.		
PR	15-JUL-1999;	99US-0144005.		
PR	16-JUL-1999;	99US-0144085.		
PR	16-JUL-1999;	99US-0144086.		
PR	19-JUL-1999;	99US-0144331.		
PR	19-JUL-1999;	99US-0144332.		
PR	19-JUL-1999;	99US-0144333.		
PR	19-JUL-1999;	99US-0144334.		
PR	19-JUL-1999;	99US-0144335.		
PR	20-JUL-1999;	99US-0144352.		
PR	20-JUL-1999;	99US-0144632.		
PR	20-JUL-1999;	99US-0144814.		
PR	21-JUL-1999;	99US-0145086.		
PR	21-JUL-1999;	99US-0145088.		
PR	22-JUL-1999;	99US-0145085.		
PR	22-JUL-1999;	99US-0145087.		
PR	22-JUL-1999;	99US-0145089.		
PR	22-JUL-1999;	99US-0145192.		
PR	23-JUL-1999;	99US-0145145.		
PR	23-JUL-1999;	99US-0145218.		
PR	23-JUL-1999;	99US-0145224.		
PR	26-JUL-1999;	99US-0145276.		
PR	27-JUL-1999;	99US-0145913.		
PR	27-JUL-1999;	99US-0145918.		
PR	27-JUL-1999;	99US-0145919.		
PR	28-JUL-1999;	99US-0145951.		
PR	02-AUG-1999;	99US-0146386.		
PR	02-AUG-1999;	99US-0146388.		
PR	02-AUG-1999;	99US-0146389.		
PR	03-AUG-1999;	99US-0147038.		
PR	04-AUG-1999;	99US-0147204.		
PR	04-AUG-1999;	99US-0147302.		
PR	05-AUG-1999;	99US-0147192.		
PR	05-AUG-1999;	99US-0147260.		
PR	06-AUG-1999;	99US-0147303.		
PR	06-AUG-1999;	99US-0147416.		
PR	09-AUG-1999;	99US-0147493.		
PR	09-AUG-1999;	99US-0147935.		
PR	10-AUG-1999;	99US-0148171.		

```
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154339.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 81.7%; Score 18.8; DB 21; Length 1152;  
Best Local Similarity 90.9%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 2 aaatccatgtaagcagcaggg 23
    |||||
Db 1088 aaatccatgtaagcaggaag 1109
```

## RESULT 5

AAC79889

ID AAC79889 standard; cdna; 1567 BP.

XX AAC79889;

AC AAC79889;

XX 09-FEB-2001 (first entry)

DT Human secreted protein encoding cdna for gene 41.

DE Human; secreted protein; cytostatic; antiarthritic; antiasthmatic;  
XX immunosuppressive; antiarteriosclerotic; antiinflammatory; nootropic;  
KW neuroprotective; antidiabetic; tranquiliser; vulnerary; antibacterial;  
KW antipsoriatic; antiarrhythmic; antirheumatic; cardiant; anti-HIV;  
KW autoimmune disorder; allergic condition; cardiovascular disorder;  
KW cancer; neurological disease; tissue repair; ss.

XX Homo sapiens.

OS WO200055176-A2.

XX 21-SEP-2000.

PD 09-MAR-2000; 2000WO-US06057.  
PF 12-MAR-1999; 99US-0124142.  
XX 11-JUN-1999; 99US-0138597.  
PR 03-DEC-1999; 99US-0168666.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

PI WPI: 2000-638176/61.

XX P-PSDB; AAB44870.

DR Novel 49 human secreted proteins useful for diagnosis, prevention and  
XX treatment of disorders including neurological, cell proliferative,  
PT cardiovascular, and autoimmune/inflammatory disorders and microbial  
PT infections -

PS Claim la; Page 356-357; 405pp; English.

XX This invention describes a novel isolated polypeptide (I) comprising an  
CC amino acid sequence at least 95 % identical to a polypeptide sequence  
CC selected from 49 polypeptides encoded by polynucleotide sequences  
CC included in American Type Culture Collection (ATCC) deposit number  
CC 203917, defined in the specification. The products of the invention have  
CC cytostatic, antiarthritic, antiasthmatic, immunosuppressive, nootropic,  
CC antiarteriosclerotic, antiinflammatory, neuroprotective, antidiabetic,  
CC tranquiliser, vulnerary, antibacterial, antipsoriatic, antiarrhythmic,  
CC antirheumatic, cardiant and anti-HIV activity. (I) or a nucleic acid (II)  
CC encoding (I) is useful for preventing, treating or ameliorating a medical  
CC condition and for diagnosing a pathological condition or susceptibility  
CC to the condition. (I) is useful for identifying a binding partner which  
CC affects the activity of the polypeptide and for identifying an activity  
CC in a biological sample. (I), (II) or an antibody (IV) specific to (I) is  
CC also useful for treating or preventing a disease, disorder or condition  
CC associated with aberrant expression of (I). Diseases treated or diagnosed  
CC include immune disorders such as autoimmune diseases, blood protein  
CC disorders, anemia, allergic reactions and conditions such as asthma,  
CC organ rejection or graft-versus-host disease, inflammation, hyper  
CC proliferative disorders, cardiovascular disorders such as arterioarterial  
CC fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ  
CC regeneration, cancer, neovascular glaucoma, diabetic retinopathy,  
CC rheumatoid arthritis, psoriasis, diseases associated with increased  
CC apoptosis that include acquired immunodeficiency syndrome (AIDS),  
CC neurological diseases such as Parkinson's disease, viral, bacterial,  
CC fungal or parasitic diseases. They are also used to repair, replace or  
CC protect tissue damage by congenital defects, to treat trauma, in surgery,  
CC including cosmetic plastic surgery, to treat fibrosis, reperfusion injury

CC or systemic cytokine damage, to stimulate chondrocyte growth, to prevent  
CC skin aging due to sunburn, to change a mammal's mental state or physical  
CC state by influencing biorhythms, cardiac rhythms, depression, memory,  
CC stress and for accelerating wound healing. (I), (II) and/or their agonist  
CC or antagonist are useful as food additives or preservatives to increase  
CC or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamin, mineral or other nutritional components. (I) is  
CC useful for screening therapeutic compounds. (II) is useful in forensic  
CC biology for detecting DNA sequences and as diagnostic probes for  
CC detecting the presence of specific mRNA in a particular cell type.  
XX  
SQ Sequence 1567 BP; 303 A; 510 C; 463 G; 291 T; 0 other;

Query Match 75.7%; Score 17.4; DB 21; Length 1567;

Best Local Similarity 94.7%; Pred. No. 71;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 atccatgtaagcagcagg 22  
||||||| |||||||||  
Db 805 atccatgtaagcagcagg 823

## RESULT 6

AAC39525

ID AAC39525 standard; DNA; 1153 BP.

XX

AC AAC39525;

XX

DT 17-OCT-2000 (first entry)

DE

Arabidopsis thaliana DNA fragment SEQ ID NO: 24952.

XX

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

PN

EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR

05-MAR-1999; 99US-0123180.

PR

09-MAR-1999; 99US-0123548.

PR

23-MAR-1999; 99US-0125788.

PR

25-MAR-1999; 99US-0126264.

PR

29-MAR-1999; 99US-0126785.

PR

01-APR-1999; 99US-0127462.

PR

06-APR-1999; 99US-0128234.

PR

08-APR-1999; 99US-0128714.

PR

16-APR-1999; 99US-0129845.

PR

19-APR-1999; 99US-0130077.

PR

21-APR-1999; 99US-0130449.

PR

23-APR-1999; 99US-0130510.

PR

23-APR-1999; 99US-0130891.

PR

28-APR-1999; 99US-0131449.

PR

30-APR-1999; 99US-0132048.

PR

30-APR-1999; 99US-0132407.

PR

04-MAY-1999; 99US-0132484.

PR

05-MAY-1999; 99US-0132485.

PR

06-MAY-1999; 99US-0132486.

PR

07-MAY-1999; 99US-0132487.

PR

11-MAY-1999; 99US-0132863.

PR

14-MAY-1999; 99US-0134218.

PR

14-MAY-1999; 99US-0134219.

PR

14-MAY-1999; 99US-0134221.

PR

14-MAY-1999; 99US-0134370.

PR

18-MAY-1999; 99US-0134768.

PR

PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.

```
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 18-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 06-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 74.8%; Score 17.2; DB 21; Length 1153;
Best Local Similarity 86.4%; Pred. No. 84;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaaacgacgagg 23
    |||||
Db 1088 aaatccatgtaaacgacgagg 1109
    |||||

RESULT 7
AAH18005
ID AAH18005 standard; cDNA; 2857 BP.
XX
AC AAH18005;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17809.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs.
XX
PS Claim 8; SEQ ID 17809; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
```



CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2857 BP; 859 A; 564 C; 576 G; 858 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 2857;  
Best Local Similarity 86.4%; Pred. No. 96;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaagcagcagg 23  
|| ||| ||||| ||||| ||  
Db 1659 aactcctgtgtaagcagcag 1680

RESULT 8  
AAI58467  
ID AAI58467 standard; cDNA: 4495 BP.

XX  
AC AAI58467;

XX  
DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 670.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX Homo sapiens.

OS  
XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX P-PSDB; AAM39511.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 670; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
XX specification.

SQ Sequence 4495 BP; 1401 A; 945 C; 1040 G; 1109 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 4495;  
Best Local Similarity 86.4%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaagcagcagg 23  
|| ||| ||||| ||||| ||  
Db 3034 aactcctgtgtaagcagcag 3055

RESULT 9

AAI60253

ID AAI60253 standard; cDNA: 4775 BP.

XX  
AC AAI60253;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4242.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX Homo sapiens.

OS  
XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX P-PSDB; AAM41097.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 4242; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC C.N.S disorders,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX  
 SQ Sequence 4775 BP; 1494 A; 999 C; 1122 G; 1160 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 4775;  
 Best Local Similarity 86.4%; Pred. No. 1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgaggg 23  
 ||| ||||| ||||| |||||  
 Db 2995 aactcctgttaaacgacgagc 3016

RESULT 10  
 AAT74989/c  
 ID AAT74989 standard; DNA; 8826 BP.

XX AAT74989;

XX 26-FEB-1998 (first entry)

DE Saccharomyces cerevisiae YHK1243 encoding CSF-1 gene.

XX Saccharomyces cerevisiae; cold sensitivity of fermentability; CSF-1;  
 KW low temperature sensitivity; yeast; bread; ethanol; fermentation; ds.  
 XX

OS Saccharomyces cerevisiae.

XX WO9724442-A1.

XX 10-JUL-1997.

XX 27-DEC-1996; 96WO-JP03862.

XX 28-DEC-1995; 95JP-0343700.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Kawasaki H, Kikuchi Y, Ouchi K, Tokai M;

XX WPI; 1997-363678/33.

DR P-PSDB; AAW22050.

XX Protein for complementing low temperature sensitivity in  
 PT fermentation - can be inactivated in yeast and used for producing  
 PT bread and ethanol

XX Claim 1; Pages 21-40; 48pp; Japanese.

XX This novel genomic DNA CSF-1 (cold sensitivity of fermentability) encodes  
 CC a protein capable of complementing a mutation showing low temperature  
 CC sensitivity in fermentation. The Saccharomyces cerevisiae YHK 1243 has an  
 CC inactivated gene encoding the new protein or has the novel DNA sequence  
 CC contained in its chromosome. The protein can be used in dough to produce  
 CC bread, and to produce ethanol.

SQ Sequence 8826 BP; 2903 A; 1551 C; 1760 G; 2612 T; 0 other;

Query Match 74.8%; Score 17.2; DB 18; Length 8826;

Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgaggg 23  
 ||| ||||| ||||| |||||  
 Db 7504 AAATTAATGTAAGCACCAGGG 7483

RESULT 11  
 AAT94548/c  
 ID AAT94548 standard; DNA; 9621 BP.

XX AAT94548;

XX 25-MAR-1998 (first entry)

XX Mutant YLR087c gene from cold sensitive yeast strain.

XX Mutant; cold sensitive; yeast; complementation; breadmaking; dough;  
 KW sugar; leavening agent; brewing; winemaking; truncation; ss.

XX Saccharomyces cerevisiae.

XX Key Location/Qualifiers

FT CDS 376..7866

FT /\*tag= a

FT /product= "mutant YLR087c protein"

XX WO9728693-A1.

XX 14-AUG-1997.

XX 07-FEB-1997; 97WO-FR00254.

XX 08-FEB-1996; 96FR-0001562.

XX (LESA ) LESAFFRE & CIE.

XX Colavizza D, Lolez A, Wadoux I;

XX WPI; 1997-414988/38.

DR P-PSDB; AAW36093.

XX New strains of bread-making yeast with low fermentative activity at  
 PT low temperature - allows production of doughs that can be stored  
 PT cold for many hours before final baking

XX Claim 10; Page 56-59; 76pp; French.

XX This is the nucleotide sequence of a mutant gene (designated YLR087c)  
 CC isolated from a cold sensitive strain of Saccharomyces cerevisiae. The  
 CC sequence was isolated from the clone YCP50-10.39 which was able to  
 CC complement the cold sensitive strain HLI3.2.30. This gene contains a  
 CC TCA to TGA mutation which encodes a truncated protein. The wild type  
 CC gene contains an open reading frame of 8874 bases encoding a protein of  
 CC 2958 amino acids, whereas this mutant gene contains an open reading frame  
 CC of 7488 bases encoding a protein of 2496 amino acids. Yeast strains  
 CC containing the YLR087c gene are used, fresh or dried, in modified  
 CC breadmaking processes, i.e. in processes where the delay between mixing  
 CC the dough and baking exceeds 6 hr. Particularly they are used to make  
 CC French-style bread (no added sugar) or breads with sugar content below  
 CC 5%. The strains can also be used to make mixed yeast/bacteria acidic  
 CC leavening agents. The strains can also be used in brewing and winemaking  
 CC as a cold-sensitive phenotype.

SQ Sequence 9621 BP; 3126 A; 1679 C; 1903 G; 2907 T; 6 other;

Query Match 74.8%; Score 17.2; DB 18; Length 9621;  
 Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgaggg 23

Db 7927 AAATTAATGTAAGCACCAGGG 7906  
||||| ||||||||| |||||

RESULT 12  
AAH29132/C  
ID AAH29132 standard; DNA; 602 BP.  
XX AC AAH29132;  
XX DT 17-JUL-2001 (first entry)  
XX DE Drosophila melanogaster essential gene fragment, SEQ ID NO: 321.  
XX KW Drosophila melanogaster; fruit fly; essential gene; screening assay;  
XX KW pesticide; crop protection; chromosome 3; ds.  
XX OS Drosophila melanogaster.  
XX PN WO200118547-A1.  
XX FN 15-MAR-2001.  
XX PD 06-SEP-2000; 2000WO-GB03444.  
XX PF 07-SEP-1999; 99GB-0021009.  
XX PR (UNIU ) UNIV GLASGOW.  
XX PA Davies RW, Kaiser K, Yang MY;  
XX PI WPI; 2001-281436/29.  
XX DR Screening assays for used for identifying compounds having a  
XX PT physiological effect on proteins identified as being essential -  
XX PT  
XX PS Claim 1; Page 340-341; 695pp; English.  
XX CC The present sequence is part of an essential gene from Drosophila  
CC melanogaster. Lack of expression of the protein encoded by this  
CC gene leads to a lethal or semi-lethal phenotype. The invention  
CC relates to 902 nucleic acid sequences from genes encoding proteins  
CC which are thought to be essential, and to a screening assay for  
CC identifying compounds which have a physiological effect on these  
CC proteins. Suitable compounds are useful as pesticides and may be used  
CC in conjunction with other pesticides and herbicides for crop  
CC protection. The gene corresponding to the present sequence is located  
CC on chromosome 3.  
XX CC  
SQ Sequence 602 BP; 159 A; 131 C; 120 G; 192 T; 0 other;

Query Match 73.0%; Score 16.8; DB 22; Length 602;  
Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgag 21  
| ||||| ||||| |||||  
Db 409 AGATCCATGTAATCAGCAG 390

RESULT 13  
AAF57408/C  
ID AAF57408 standard; DNA; 1752 BP.  
XX AC AAF57408;  
XX DT 11-JUN-2001 (first entry)  
XX DE Human p95 protein encoding DNA.  
XX KW Rab; Rho; GTPase; pharmaceutical; cancer; anti-HIV; tuberculostatic;  
KW protozoacide; atidiabetic; nootropic; neuroprotective; dermatological;

KW antipsoriatic; antiinflammatory; antiallergic; antipyretic; cytostatic;  
KW antibacterial; gynecological; antiatherosclerotic; gene therapy; human;  
KW p100; p95; p60; p45; p25; endocytic trafficking; GTPase effector; ds.  
XX OS Homo sapiens.  
XX Key Location/Qualifiers  
FH 1..1752  
FT /\*tag= a  
FT /product= "p95"  
XX WO200120022-A1.  
XX PN 22-MAR-2001.  
XX PD 18-SEP-2000; 2000WO-EP09130.  
XX PF 16-SEP-1999; 99EP-0118385.  
XX PR (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX PA Nielsen E, Christophoridis S, Murphy C, Zerial M, De Renzis S;  
XX PI WPI; 2001-257888/26.  
XX DR P-PSDB; AAB62180.  
XX PT Use of effectors of GTPase as target in a in vitro/vivo assay for  
XX PT detecting substances for prophylaxis, treatment of cancer, cell  
XX PT migration disorders, e.g. Alzheimer's, infectious diseases, diabetes,  
XX PT atherosclerosis  
XX PS Claim 10; Page 70-71; 76pp; English.  
XX CC The invention relates to the use of effectors/regulators for Rab and Rho  
CC GTPases in in vitro or in vivo assays. The method is useful for detecting  
CC substances useful as pharmaceutical agents for the prophylaxis or  
CC treatment of cancer and other proliferative, invasive or cell migration  
CC disorders such as endometriosis, atherosclerosis, inflammatory and  
CC allergic diseases, infectious diseases, diabetes, Alzheimer's disease and  
CC skin repair diseases such as psoriasis. The infectious diseases include  
CC AIDS, tuberculosis, pseudotuberculosis, cholera, gastroenteritis, enteric  
CC fever, malaria, typhus, diseases caused by pathogens such as Listeria,  
CC Mycobacterium, Staphylococcus, Toxoplasma, Trypanosoma, Salmonella,  
CC Legionella, Leishmania, Coxiella, Shigella, Yersinia, Neisseria, Vibrio,  
CC Bartonella. The cancer includes benign tumor, malignant lung carcinoma,  
CC leukemia, glioma or a neuroblastoma, in particular lung carcinoma,  
CC osteosarcoma, lymphoma, breast, bile, intestine, kidney, ovary, stomach,  
CC brain, prostate, liver and every tumor that invades other tissues and  
CC organs distinct from its site of origin. The assay is highly sensitive  
CC and advantageous in the selectivity of the targets. The present sequence  
CC represents the DNA encoding a human p95 protein. A multiprotein complex  
CC including p100, p95, p60, p45, p25 (sequences AAF57404-408) acts as a  
CC GTPase effector/regulator and has the function of endocytic trafficking.  
XX CC  
SQ Sequence 1752 BP; 496 A; 425 C; 444 G; 387 T; 0 other;

Query Match 73.0%; Score 16.8; DB 22; Length 1752;  
Best Local Similarity 90.0%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaatccatgtaaacgacga 20  
| ||||| ||||| |||||  
Db 712 GAAATCCATGCAGACGACGA 693

RESULT 14  
AAT60073  
ID AAT60073 standard; DNA; 4388 BP.  
XX AC AAT60073;  
XX DT 13-MAY-1997 (first entry)

XX Selective marker gene.  
DE  
XX  
KW Selective marker; enzyme: biosynthetic system; leucine synthesis;  
KW Rhizomucor pusillus; protein production; ss.  
XX  
OS Synthetic.  
XX  
XX JP09009971-A.  
PN  
PD 14-JAN-1997.  
XX  
XX 30-JUN-1995; 95JP-0166115.  
PF  
XX 30-JUN-1995; 95JP-0166115.  
PR  
XX (MEIT ) MEITO SANGYO KK.  
PA  
XX  
XX WPI; 1997-126428/12.  
DR  
XX Novel selective marker gene - used in a transformation system of  
PT Rhizomucor pusillus for protein production  
XX  
XX Claim 1; Page 9-10; 12pp; Japanese.  
PS  
XX This sequence represents the selective marker gene of the invention. This  
CC gene encodes an enzyme relating to the biosynthetic system of leucine  
CC synthesis. This sequence can be used in a recombinant vector for the  
CC transformation of Rhizomucor pusillus. The recombinant vector also  
CC includes another DNA sequence encoding a protein. The plasmid and the  
CC transformed Rhizomucor pusillus can then be used for the efficient  
CC production of useful proteins and peptides, which are encoded by the  
CC second DNA sequence in the vector.  
XX  
XX Sequence 4388 BP; 1144 A; 1094 C; 1077 G; 1073 T; 0 other;  
SQ

Query Match 73.0%; Score 16.8; DB 18; Length 4388;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 gaataccatgttaagcagca 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 611 gaatgcataaaagcagca 630

RESULT 15  
AAQ60834  
ID AAQ60834 standard; DNA; 281 BP.  
XX  
XX AC AAQ60834;  
XX  
DT 16-MAR-1994 (first entry)  
XX  
XX Human brain Expressed Sequence Tag EST00937.  
DE  
XX  
XX Gene transcription product; genetic markers; tagging; in vivo;  
KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO9316178-A.  
PN  
XX  
PD 19-AUG-1993.  
XX  
PF 12-FEB-1993; 93WO-US01294.  
XX  
PR 12-FEB-1992; 92US-0837195.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
XX  
PI Adams MD, Moreno RF, Venter CJ;  
XX

DR WPI; 1993-272882/34.  
XX  
XX Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
PT of most human genes  
XX  
XX Example 4; Page 398; 500pp; English.  
PS  
XX The Expressed Sequence Tag was isolated from a human brain cDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue  
CC type, and for prep. of antisense sequences, probes and constructs.  
CC EST00937 has a "poor" coding probability as evaluated using the  
CC coding-region prediction program CRM. See also AAQ59041-Q61440.  
XX  
SQ Sequence 281 BP; 45 A; 81 C; 91 G; 61 T; 3 other;  
Query Match 72.2%; Score 16.6; DB 14; Length 281;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 gaataccatgttaagcagcaggg 23  
| | | | | | | | | | | | | | | | | | | | | |  
Db 86 gaagtgaatgcaagcagcaggg 108

Search completed: February 25, 2002, 18:17:40  
Job time: 16698 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:49 ; Search time 301.6 Seconds  
(without alignments)  
17.271 Million cell updates/sec

Title: US-09-698-903B-11

Perfect score: 23

Sequence: 1 gaatccatgtaaagcagcagg 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.2	74.8	8874	US-08-894-344C-1	Sequence 1, Appli
C 2	17.2	74.8	9621	US-09-125-028-1	Sequence 1, Appli
C 3	16.6	72.2	2460	US-08-964-127-1	Sequence 1, Appli
C 4	16.6	72.2	9751	US-09-238-303-7	Sequence 7, Appli
C 5	16.2	70.4	27	US-08-832-399-9	Sequence 9, Appli
C 6	16.2	70.4	27	US-09-372-498-9	Sequence 9, Appli
C 7	16.2	70.4	717	US-08-706-702-25	Sequence 25, Appli
C 8	16.2	70.4	717	US-08-706-706-25	Sequence 25, Appli
C 9	16.2	70.4	1351	US-09-002-114-2	Sequence 2, Appli
C 10	16.2	70.4	1608	US-08-899-371-2	Sequence 2, Appli
C 11	16.2	70.4	1613	US-08-899-371-1	Sequence 1, Appli
C 12	16.2	70.4	5183	US-09-039-555B-18	Sequence 18, Appli
C 13	16.2	70.4	5243	US-08-414-335-2	Sequence 2, Appli
C 14	16.2	70.4	16442	US-08-781-891-208	Sequence 208, App
C 15	15.8	68.7	199	US-08-250-346-21	Sequence 21, Appl
C 16	15.8	68.7	394	US-07-885-089B-15	Sequence 15, Appl
C 17	15.8	68.7	394	US-07-885-089B-17	Sequence 17, Appl
C 18	15.8	68.7	3410	US-09-020-956-110	Sequence 110, App
C 19	15.8	68.7	3410	US-09-030-607-110	Sequence 110, App
C 20	15.8	68.7	3466	US-08-468-036-38	Sequence 38, Appl
C 21	15.8	68.7	3466	US-08-376-843-38	Sequence 38, Appl
C 22	15.6	67.8	1098	US-08-948-616-6	Sequence 6, Appli
C 23	15.6	67.8	1098	US-09-193-510-6	Sequence 6, Appli
C 24	15.6	67.8	1098	US-09-368-402-6	Sequence 6, Appli
C 25	15.6	67.8	1488	US-09-171-969-1	Sequence 1, Appli
C 26	15.4	67.0	2266	US-09-213-767-1	Sequence 1, Appli
C 27	15.4	67.0	152331	US-09-128-155-16	Sequence 16, Appli

C 28	15.4	67.0	176373	3	US-09-128-155-17	Sequence 17, Appl
C 29	15.2	66.1	681	5	PCT-US94-04361-22	Sequence 22, Appl
C 30	15.2	66.1	1338	2	US-08-484-993B-7	Sequence 7, Appli
C 31	15.2	66.1	1338	2	US-08-484-158B-7	Sequence 7, Appli
C 32	15.2	66.1	1338	2	US-08-484-596A-7	Sequence 7, Appli
C 33	15.2	66.1	1338	2	US-08-480-150A-7	Sequence 7, Appli
C 34	15.2	66.1	1338	3	US-08-458-731-7	Sequence 7, Appli
C 35	15.2	66.1	1338	3	US-08-149-223A-7	Sequence 7, Appli
C 36	15.2	66.1	2975	4	US-08-368-281-1	Sequence 3, Appli
C 37	15.2	66.1	5107	4	US-08-910-647-3	Sequence 3, Appli
C 38	15.2	66.1	18994	1	US-08-459-586-4	Sequence 4, Appli
C 39	15.2	66.1	18994	2	US-08-282-696-4	Sequence 4, Appli
C 40	15	65.2	35	1	US-08-094-534-18	Sequence 18, Appl
C 41	15	65.2	35	2	US-08-581-543-18	Sequence 18, Appl
C 42	15	65.2	35	5	PCT-US94-08000-18	Sequence 18, Appl
C 43	15	65.2	39	1	US-08-094-534-17	Sequence 17, Appl
C 44	15	65.2	39	2	US-08-581-543-17	Sequence 17, Appl
C 45	15	65.2	39	5	PCT-US94-08000-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1  
US-08-894-344C-1/c  
; Sequence 1, Application US/08894344C  
; Patent No. 6172196  
; GENERAL INFORMATION:  
; APPLICANT: KAWASAKI, Hideki  
; APPLICANT: TOKAI, Masaya  
; APPLICANT: KIKUCHI, Yasuhiro  
; APPLICANT: OUCHI, Koza  
; TITLE OF INVENTION: DNA ENCODING PROTEIN COMPLEMENTING  
; TITLE OF INVENTION: YEAST  
; TITLE OF INVENTION: LOW TEMPERATURE-SENSITIVE FERMENTABILITY  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FITZPATRICK, CELLA, HARPER & SCINTO  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10112-3801  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette - 3.50 inch, 1440 Kb storage.  
; COMPUTER: IBM PS/V  
; OPERATING SYSTEM: MS-DOS Ver3.30  
; SOFTWARE: PATENT AID Ver1.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/894,344C  
; FILING DATE: 15-AUGUST-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP343700/95  
; FILING DATE: 28-DECEMBER-1995  
; APPLICATION NUMBER: PCT/JP96/03862  
; FILING DATE: 27-DECEMBER-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perry, Lawrence S.  
; REGISTRATION NUMBER: 31865  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-218-2100  
; TELEFAX: 212-218-2200  
; INFORMATION FOR SEQ ID NO: 1 :  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8874 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE: *Saccharomyces cerevisiae*  
; ORGANISM: *Saccharomyces cerevisiae*  
; STRAIN: X2180-1B  
; FEATURE:

```
; NAME/KEY: CDS
; LOCATION: 1 to 8874
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 1291 to 1296
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 4388 to 4393
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 5927 to 5032
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 7675 to 7680
; IDENTIFICATION METHOD: S
; US-08-894-344C-1
```

```
Query Match 74.8%; Score 17.2; DB 4; Length 8874;
Best Local Similarity 86.4%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 aaatccatgtaaacgacgagg 23
||||| ||||| ||||| |||||
Db 7552 AAATTAATGTAAGCACCAGG 7531
```

```
RESULT 2
US-09-125-028-1/c
; Sequence 1, Application US/09125028A
; Patent No. 6190707
; GENERAL INFORMATION:
; APPLICANT: WADOUX, Isabelle
; APPLICANT: COLAVIZZA, Didier
; TITLE OF INVENTION: NOVEL COLD-SENSITIVE BREAD-MAKING YEASTS
; FILE REFERENCE: levure sensible froid
; CURRENT APPLICATION NUMBER: US/09/125,028A
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: PCT/FR97/00254
; EARLIER FILING DATE: 1997-02-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 9621
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9318)
; OTHER INFORMATION: Use of n signifies any of g, a, c or t
US-09-125-028-1
```

```
Query Match 74.8%; Score 17.2; DB 4; Length 9621;
Best Local Similarity 86.4%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 aaatccatgtaaacgacgagg 23
||||| ||||| ||||| |||||
Db 7927 AAATTAATGTAAGCACCAGG 7906
```

```
RESULT 3
US-08-964-127-1/c
; Sequence 1, Application US/08964127
; Patent No. 6277565
; APPLICANT: Grandearl, Andrew David John
```

```
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,127
; FILING DATE: 06-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crews, Ph.D., L. Lee
; REGISTRATION NUMBER: P-43,567
; REFERENCE/DOCKET NUMBER: 07334/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 498....2057
; US-08-964-127-1
```

```
Query Match 72.2%; Score 16.6; DB 4; Length 2460;
Best Local Similarity 82.6%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 gaaatccatgtaaacgacgagg 23
||||| ||||| ||||| |||||
Db 2242 GAAGTGAATGCAAGCAGCAGG 2220
```

```
RESULT 4
US-09-238-303-7
; Sequence 7, Application US/09238303B
; Patent No. 6284253
; GENERAL INFORMATION:
; APPLICANT: Barr, Margaret C.
; TITLE OF INVENTION: No. 6284253el feline Immunodeficiency Virus Nucleotide Sequence
; FILE REFERENCE: 18617.0059
; CURRENT APPLICATION NUMBER: US/09/238,303B
; CURRENT FILING DATE: 1999-01-28
; EARLIER APPLICATION NUMBER: US 60/072,927
; EARLIER FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 7
; LENGTH: 9751
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of
; OTHER INFORMATION: a Pallas's cat feline immunodeficiency virus
US-09-238-303-7
```



; GENERAL INFORMATION:  
 ; APPLICANT: Derk J. Bergsma  
 ; APPLICANT: Usman Shabon  
 ; TITLE OF INVENTION: NOVEL F

; GENERAL INFORMATION:  
 ; APPLICANT: Derk J. Bergsma  
 ; APPLICANT: Usman Shabon  
 ; TITLE OF INVENTION: NOVEL F

TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:

; LENGTH: 717 base pairs

RESULT 10  
US-08-899-371-2

```
; Sequence 2, Application US/08899371
; Patent No. 5962227
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Ronald P.
; APPLICANT: Andree, Karl B.
; APPLICANT: Antonio, Dolores B.
; TITLE OF INVENTION: A DNA-Based Diagnostic Test for
; TITLE OF INVENTION: Detecting Myxobolus, the Cause of Salmonid Whirling
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsens and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,371
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/022,734
; FILING DATE: 26-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-081310US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..1608
; OTHER INFORMATION: /note= "18S rRNA gene of Myxobolus
; insidiosus"
;
; US-08-899-371-2
;
; Query Match 70.4%; Score 16.2; DB 2; Length 1608;
; Best Local Similarity 85.7%; Pred. No. 59;
; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; Qy 2 aaatccatgtaagcagcagg 22
; | ||||| || |||||
; Db 312 ACATCCATGGAAGCAGCAGG 332
;
; RESULT 11
; US-08-899-371-1
; Sequence 1, Application US/08899371
; Patent No. 5962227
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Ronald P.
; APPLICANT: Andree, Karl B.
; APPLICANT: Antonio, Dolores B.
; TITLE OF INVENTION: A DNA-Based Diagnostic Test for
; TITLE OF INVENTION: Detecting Myxobolus, the Cause of Salmonid Whirling
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Townsens and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,371
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/022,734
; FILING DATE: 26-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-081310US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..1613
; OTHER INFORMATION: /note= "18S rRNA gene of Myxobolus
; cerebralis"
;
; US-08-899-371-1
;
; Query Match 70.4%; Score 16.2; DB 2; Length 1613;
; Best Local Similarity 85.7%; Pred. No. 59;
; Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; Qy 2 aaatccatgtaagcagcagg 22
; | ||||| || |||||
; Db 315 ACATCCATGGAAGCAGCAGG 335
;
; RESULT 12
; US-09-039-555B-18/C
; Sequence 18, Application US/09039555B
; Patent No. 6033856
; GENERAL INFORMATION:
; APPLICANT: Koerner, Kathrin
; APPLICANT: Mueller, Rolf
; APPLICANT: Sadlacek, Hans-Harald
; TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

; APPLICATION NUMBER: US/09/039,555B  
; FILING DATE: 16-MAR-1998  
; CLASSIFICATION: 514  
; PRIOR APPLICATION NUMBER: DE 19710643.9  
; FILING DATE: 14-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 016779/0131  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5183 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-09-039-555B-18

Query Match 70.4%; Score 16.2; DB 3; Length 5183;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgag 22  
||||| ||| ||| ||| |||  
Db 741 AAATCCAGCTATAGCAGCAGG 721

RESULT 13  
US-08-414-335-2  
; Sequence 2, Application US/08414335  
; Patent No. 5907078  
; GENERAL INFORMATION:  
; APPLICANT: GREENBERT, No. 5907078man M  
; APPLICANT: MATUSIK, Robert J  
; APPLICANT: ROSEN, Jeffrey M  
; TITLE OF INVENTION: TRANSGENIC MOUSE MODEL FOR PROSTATE  
; TITLE OF INVENTION: CANCER  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare Ltd.  
; STREET: 1203 Crystal Plaza Bldg. I, 2001 Jefferson  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: United States of America  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/414,335  
; FILING DATE: 31-MAR-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FALLOW, Charles W  
; REGISTRATION NUMBER: 28,946  
; REFERENCE/DOCKET NUMBER: 1027  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5243 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single

; TOPOLOGY: linear  
US-08-414-335-2

Query Match 70.4%; Score 16.2; DB 2; Length 5243;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgag 22  
||||| ||| ||| ||| |||  
Db 4443 AAATCCAGCTATAGCAGCAGG 4463

RESULT 14  
US-08-781-891-208/c  
; Sequence 208, Application US/08781891  
; Patent No. 6090620  
; GENERAL INFORMATION:  
; APPLICANT: Fu, Ying-Hui  
; APPLICANT: Yu, Chang-En  
; APPLICANT: Oshima, Junko  
; APPLICANT: Mulligan, John T.  
; APPLICANT: Schellenberg, Gerald D.  
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
; TITLE OF INVENTION: WERNER'S SYNDROME  
; NUMBER OF SEQUENCES: 209  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,891  
; FILING DATE: 27-DEC-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6090620tenburg Ph.D., Carol  
; REGISTRATION NUMBER: 39,317  
; REFERENCE/DOCKET NUMBER: 240052.419  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 208:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 16442 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-781-891-208

Query Match 70.4%; Score 16.2; DB 3; Length 16442;  
Best Local Similarity 85.7%; Pred. No. 94;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaatccatgtaaacgacgag 21  
||||| ||| ||| ||| |||  
Db 12904 GAAATCCATGTAAGTCACAG 12884

RESULT 15  
US-08-250-346-21/c  
; Sequence 21, Application US/08250346  
; Patent No. 5939255  
; GENERAL INFORMATION:  
; APPLICANT: ANAND, Rakesh

APPLICANT: MARKHAM, Alexander F  
APPLICANT: SMITH, John C  
APPLICANT: ANWAR, Rashida  
APPLICANT: RILEY, John H  
APPLICANT: OGILVIE, Donald J  
APPLICANT: ELVIN, Paul  
TITLE OF INVENTION: DIAGNOSTIC METHOD  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/250,346  
FILING DATE: 27-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/115,675  
FILING DATE: 02-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/879,117  
FILING DATE: 04-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/578,616  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 8920211.3  
FILING DATE: 07-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: BIRD, Donald J.  
REGISTRATION NUMBER: 25,323  
REFERENCE/DOCKET NUMBER: 205356/PHM35417/USC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 199 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-250-346-21

Query Match 68.7%; Score 15.8; DB 2; Length 199;  
Best Local Similarity 89.5%; Pred. No. 60;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 tccatgtaaagcagcagg 23  
| | | | |  
Db 178 TCCAGGAAAGCAGCAGG 160

Search completed: February 25, 2002, 18:05:51  
Job time: 18599 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:13 ; Search time 8261.74 seconds  
(without alignments)  
29.915 Million cell updates/sec

Title: US-09-698-903B-11

Perfect score: 23  
Sequence: 1 gaataccatgtaaacgacgagg 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estfun:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estom:\*
  - 5: em\_estpl:\*
  - 6: em\_estba:\*
  - 7: em\_estro:\*
  - 8: em\_estov:\*
  - 9: em\_hic:\*
  - 10: gb\_estl:\*
  - 11: gb\_est2:\*
  - 12: gb\_hic:\*
  - 13: gb\_gss:\*
  - 14: em\_gss\_fun:\*
  - 15: em\_gss\_hum:\*
  - 16: em\_gss\_inv:\*
  - 17: em\_gss\_pln:\*
  - 18: em\_gss\_pro:\*
  - 19: em\_gss\_rod:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.8	81.7	619	13	FR0032239
2	18.8	81.7	916	13	AL028607 Fugu rubr
3	18.8	81.7	1007	13	AL315856 Tetraodon
C 4	18.4	80.0	342	13	AL345585 Tetraodon
C 5	18.4	80.0	522	13	AZ842531
6	18.4	80.0	618	13	AQ972989 Fugu rubr
C 7	18.2	79.1	205	10	FR0034498
C 8	18.2	79.1	287	10	AV119745
C 9	18.2	79.1	444	10	AV119723
10	18.2	79.1	455	13	AI243356
C 11	18.2	79.1	481	10	AZ628185
C 12	18.2	79.1	533	13	BE019857
					AZ484979

C 13	18.2	79.1	750	13	AZ750976
14	18.2	79.1	809	11	BF027105
C 15	18.2	79.1	875	11	BG118586
16	18.2	79.1	951	13	CNS0528P
C 17	17.8	77.4	193	13	AZ658541
18	17.8	77.4	287	10	BE579037
C 19	17.8	77.4	525	10	AV395870
20	17.8	77.4	563	10	AV390896
C 21	17.8	77.4	564	10	AV387535
22	17.8	77.4	582	13	AZ107460
C 23	17.8	77.4	595	10	AA394343
24	17.8	77.4	618	10	AM661616
C 25	17.8	77.4	618	13	AZ558735
26	17.8	77.4	626	10	BE337492
C 27	17.8	77.4	630	10	BE238191
28	17.8	77.4	657	10	AW054024
C 29	17.8	77.4	700	11	EG846208
30	17.8	77.4	723	11	EG845360
C 31	17.8	77.4	772	10	AU080746
32	17.4	75.7	221	10	AL507374
C 33	17.4	75.7	364	10	BE067433
34	17.4	75.7	604	13	AQ372694
C 35	17.4	75.7	857	13	CNS0660P
36	17.4	75.7	975	11	BF348682
C 37	17.4	75.7	1011	13	CNS0208G
38	17.2	74.8	158	11	BI268414
C 39	17.2	74.8	226	10	AA298276
40	17.2	74.8	271	10	AW357388
C 41	17.2	74.8	312	10	AI905219
42	17.2	74.8	312	11	H20500
C 43	17.2	74.8	332	10	BB251925
44	17.2	74.8	340	10	AA012198
C 45	17.2	74.8	357	11	W63225

## ALIGNMENTS

RESULT 1

FR0032239					
LOCUS	FR0032239	619 bp	DNA	GSS	27-JUN-1998
DEFINITION	Fugu rubripes GSS sequence, clone 137L16aC10, genomic survey sequence.				
ACCESSION	AL028607				
VERSION	AL028607.1	GI:3270721			
KEYWORDS	GSS: genome survey sequence.				
SOURCE	Takifugu rubripes.				
ORGANISM	Takifugu rubripes				
REFERENCE	1 (bases 1 to 619)				
AUTHORS	Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrantia,Y., Williams,G. and Brenner,S.				
	Direct Submission				
	Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK. Email: bhoehelp@hmp.mrc.ac.uk				
	Vector: pBluescript II KS V.type: phagemid PRIMER: KS				

One pass dye-terminator sequencing of cosmid cloned genomic sequence.

FEATURES  
source  
1. 619  
/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid.137L16"  
/clone="137L16aC10"

BASE COUNT  
ORIGIN  
131 a 178 c 145 g 116 t 49 others

[illegible]



```

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
          University of Utah Genome Center
          Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
          84112, USA
          Tel: 801 585 5606
          Fax: 801 585 7177
          Email: gdunn@genetics.utah.edu
          Insert Length: 10000 Std Error: 0.00
          Plate: 0141 row: J column: 03
          Seq primer: CGTGTAAAGCAGCGCCAGT
          Class: plasmid ends
          High quality sequence stop: 342.
          Location/Qualifiers
            1. 342
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="UUGC2M014J03"
              /clone_lib="Mouse 10kb plasmid UUGCLM library"
              /sex="Male"
              /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
              /note="Vector: PWD42nv; Purified genomic DNA from M.
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
              adaptor DNA was purified and size-selected for a 9.5 to
              10.5 kb range using preparative agarose gel
              electrophoresis. Vector DNA was prepared from a derivative
              of PWD42 (gil4732114[gb|AF129072.1]), a copy-number
              inducible derivative of plasmid RL. The vector was ligated
              with Adaptors complementary to the insert Adaptors and
              purified. The sheared, adaptor mouse DNA was annealed to
              adaptor vector DNA, and transformed into
              chemically-competent E. coli XL10-Gold (Stratagene) cells
              and selected for ampicillin resistance."
            50 a 84 c 82 g 126 t
            ORIGIN

          Query Match 80.0%; Score 18.4; DB 13; Length 342;
          Best Local Similarity 95.0%; Pred. No. 3.8e+02;
          Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

          Qy 4 atccatgtaaagcagcaggg 23
          | ||||| ||||| ||||| |||||
          Db 174 ACCCATGTAAAGCAGCAGGG 155

          RESULT 5
          LOCUS AO972989/c
          DEFINITION RPCI-23-317G14.TV RPCI-23 Mus musculus genomic clone RPCI-23-317G14
          , DNA sequence.
          ACCESSION AO972989
          VERSION AO972989.1 GI:6803442
          KEYWORDS GSS.
          SOURCE house mouse.
          ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          1 (bases 1 to 522)
          REFERENCE Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret
          , B., Levins, W., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
          and Fraser, C.M.
          Mouse BAC End Sequences from Library RPCI-23
          Unpublished (1999)

          Other_GSSs: RPCI-23-317G14.TJ
          Contact: Shaying Zhao
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: szhao@tigr.org
          Clones are derived from the mouse BAC library RPCI-23. For BAC
          library availability, please contact Pieter de Jong
          (pieter@dejong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
          or from Resea ch Genetics (info@resgen.com). BAC end page:
          http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
          Plate: 317 row: G column: 14
          Seq primer: T7
          Class: BAC ends.
          Location/Qualifiers
            1. 522
              /organism="Mus musculus"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="RPCI-23-317G14"
              /clone_lib="RPCI-23"
              /sex="Female"
              /lab_host="DH10B"
              /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site.1:
              EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
              brain genomic DNA was isolated and partially digested
              with a combination of EcoRI and EcoRI Methylase. Size
              selected DNA was cloned into the pBACE3.6 vector at the
              EcoRI sites. The ligation products were transformed into
              DH10B electrocompetent cells (BRL Life Technologies). "
            90 a 159 c 106 g 164 t 3 others
            ORIGIN

          Query Match 80.0%; Score 18.4; DB 13; Length 522;
          Best Local Similarity 95.0%; Pred. No. 4e+02;
          Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

          Qy 4 atccatgtaaagcagcaggg 23
          | ||||| ||||| ||||| |||||
          Db 394 ACCCATGTAAAGCAGCAGGG 375

          RESULT 6
          LOCUS FR0034498
          DEFINITION Fugu rubripes GSS sequence, clone 199F09af9, genomic survey
          sequence.
          ACCESSION AL030865
          VERSION AL030865.1 GI:3272979
          KEYWORDS GSS; genome survey sequence.
          SOURCE Takifugu rubripes
          ORGANISM Takifugu rubripes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
          Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
          Tetraodontidae; Takifugu.
          1 (bases 1 to 618)
          REFERENCE Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrantia, Y.,
          Williams, G. and Brenner, S.
          Direct Submission
          Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
          Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:
          biohelp@hmp.mrc.ac.uk
          Vector: pBluescript II KS
          V_type: phagemid
          PRIMER: KS
          DESCR:
          One pass dye-terminator sequencing of cosmid cloned genomic
          sequence.

```

FEATURES  
source

Location/Qualifiers  
1..618

/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid 199F09"  
/clone="199F09aF9"

BASE COUNT 109 a 183 c 154 g 141 t 31 others  
ORIGIN

## Query Match

Best Local Similarity 80.0%; Score 18.4; DB 13; Length 618;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaagcagcagg 23

Db 362 ACATCCNNGTAAAGCAGCAGG 383

## RESULT 7

AV119745/c

LOCUS AV119745 205 bp mRNA 30-JUN-1999  
DEFINITION AV119745 Mus musculus C57BL/6J 10-day embryo Mus musculus cDNA  
clone 2610305L08, mRNA sequence.

ACCESSION AV119745

VERSION AV119745.1 GI:5301896

KEYWORDS EST.

SOURCE house mouse.

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS

1 (bases 1 to 205)  
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara  
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,  
Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,  
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara  
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,  
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,  
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

## TITLE

RIKEN Mouse ESTs

JOURNAL Unpublished (1999)

COMMENT Contact: Chie Owa

Genome Science Laboratory

RIKEN 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp  
Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length cDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.

## FEATURES

source

Location/Qualifiers

1..205  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="2610305L08"  
/clone\_lib="Mus musculus C57BL/6J 10-day embryo"  
/sex="mixed"  
/dev\_stage="10-day embryo"

BASE COUNT 42 a 36 c 37 g 90 t

## ORIGIN

## Query Match

Best Local Similarity 79.1%; Score 18.2; DB 10; Length 205;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

1 gaaatccatgtaaagcagcagg 23

Db 192 GAAATCCAAGTAAACAGAGG 170

## RESULT 8

AV119723/c

LOCUS AV119723 287 bp mRNA

DEFINITION AV119723 Mus musculus C57BL/6J 10-day embryo Mus musculus cDNA

clone 2610305J08, mRNA sequence.

ACCESSION AV119723

VERSION AV119723.1 GI:5301874

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 287)

AUTHORS

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara  
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,  
Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,  
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara  
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,  
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,  
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

RIKEN Mouse ESTs

JOURNAL Unpublished (1999)

COMMENT Contact: Chie Owa

Genome Science Laboratory

RIKEN 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

Location/Qualifiers

1..287

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="2610305J08"

/clone\_lib="Mus musculus C57BL/6J 10-day embryo"

/sex="mixed"

/dev\_stage="10-day embryo"

BASE COUNT 66 a 57 c 52 g 112 t

ORIGIN

Query Match

Best Local Similarity 79.1%; Score 18.2; DB 10; Length 287;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaaatccatgtaaagcagcagg 23

Db 274 GAAATCCAAGTAAACAGAGG 252

## RESULT 9

AI243356/c

LOCUS AI243356 444 bp mRNA

DEFINITION AI243356 Homo sapiens cDNA clone

IMAGE:1846226 3', mRNA sequence.

ACCESSION AI243356

VERSION AI243356.1 GI:3838753

KEYWORDS EST.

SOURCE human.

**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** NCBI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
**AUTHORS** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
**TITLE** Tumor Gene Index  
Unpublished (1997)  
**JOURNAL** Contact: Robert Strausberg, Ph.D.  
**COMMENT** Email: cgaaps@rmail.nih.gov  
This clone is available royalty-free through LLNL : contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1011 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 424.

**FEATURES**

Location/Qualifiers	1..444
/organism=	"Homo sapiens"
/db_xref=taxon:	9606"
/clone_lib=	"IMAGE:1846226"
/clone_host=	"Soares_NFL_T_GBC_S1"
/lab_host=	"DH10B"
/notes=Organ:	pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
	Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Tatina Bonaldo.
BASE COUNT	138 a 89 c 97 g 119 t 1 others
ORIGIN	

Query Match 79.1%; Score 18.2; DB 10; Length 444;  
Best Local Similarity 87.0%; Pred.No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**QY** 1 gaatccatgtaaaagcagcgagg 23  
|||||  
**Dd** 383 GCAACCAATGTAAAGCAGCAGGG 361

**RESULT** 10  
**AZ628185**  
**LOCUS** AZ628185 455 bp DNA GSS 13-DEC-2000  
**DEFINITION** IM0480H08F Mouse 10kb plasmid UUGCIM library Mus musculus genomic  
clone UUGCIM0480H08 F, DNA sequence.  
**ACCESSION** AZ628185  
**VERSION** AZ628185.1 GI:11750375  
**KEYWORDS** GSS.  
**SOURCE** house mouse.  
**ORGANISM** Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE** Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
**AUTHORS** Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.,  
and Wright,D.;Weiss,R.  
**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)

**JOURNAL** Contact: Robert B. Weiss  
**COMMENT** University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCT, UT  
84112, USA  
Tel.: 801 585 5606

image.llnl.gov/image/html/iresources.shtml

Trace considered overall poor quality  
Seq primer: -40RP from Gibco  
High quality sequence stop: 1.

#### FEATURES

Location/Qualifiers  
1..481  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3029912"  
/clone\_lib="NIH\_MGC\_9"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

#### BASE COUNT

117 a 127 c 138 g 99 t

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 481;  
Best Local Similarity 87.0%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgagg 23  
|||||  
Db 259 GAAGTCCATGTAGAGCAACAGGG 237

#### RESULT 12

AZ484979/c  
LOCUS 533 bp DNA GSS 05-OCT-2000  
DEFINITION 1M0311P16R mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0311P16 R, DNA sequence.  
ACCESSION AZ484979  
VERSION 1 GI:10650348  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 533)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 306, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0311 row: P column: 16  
Seq primer: CACACGAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 533.  
Location/Qualifiers  
1..533  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0311P16"

#### FEATURES

Location/Qualifiers  
1..533  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0311P16"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, P-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

#### BASE COUNT

139 a 119 c 103 g 172 t

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 13; Length 533;  
Best Local Similarity 87.0%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgagg 23  
|||||  
Db 146 GAAGTCTATGTAAAGCAGCAGAG 124

#### RESULT 13

AZ750976/c  
LOCUS 750 bp DNA GSS 25-JAN-2001  
DEFINITION RPCI-24-128J23-TV RPCI-24 Mus musculus genomic clone RPCI-24-128J23  
, DNA sequence.  
ACCESSION AZ750976  
VERSION 1 GI:12536135  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 750)  
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.  
TITLE Mouse BAC End Sequences from Library RPCI-24  
JOURNAL Unpublished (1999)  
COMMENT Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 128 row: J column: 23  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1..750  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

/clone="RPCI-24-128J23"		/clone_lib="RPCI-24"	
/sex="Male"		/cell_type="Spleen/Brain"	
/note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."			
BASE COUNT	203 a	182 c	182 g
ORIGIN	183 t		
Query Match			
Best Local Similarity	79.1%; Score 18.2; DB 13; Length 750;		
Matches	20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		
OY	1	gaatccatgtaaacgacgagg	23
DB	533	CGAGCCATGTAAAGCCGACGG	511
RESULT 14			
BF027105	809 bp mRNA EST 10-OCT-2000		
LOCUS	601670224F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3953500 5',		
DEFINITION	mRNA sequence.		
ACCESSION	BF027105		
VERSION	BF027105.1 GI:10734817		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 809)		
TITLE	NTH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-r@mail.nih.gov		
	Tissue Procurement: ATCC/DCTD/DTP		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov		
	Plate: L1CM824 row: 0 column: 05		
	High quality sequence start: 32		
	High quality sequence stop: 801.		
FEATURES	Location/Qualifiers		
source	1..809		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3953500"		
	/clone_lib="NIH_MGC_20"		
	/tissue_type="melanotic melanoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		
BASE COUNT	251 a	167 c	207 g
ORIGIN	184 t		
Query Match			
Best Local Similarity	79.1%; Score 18.2; DB 11; Length 809;		
Matches	20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		

---



```

BASE COUNT      6 a      4 c      4 g      7 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 21; DB 6; Length 21;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21
|||||
Db 1 GCTGGACTATAACTTGAC 21

RESULT 2
AXI172451
LOCUS AXI172451 21 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 12 from Patent WO0141558.
ACCESSION AXI172451
VERSION AXI172451.1 GI:14597563
KEYWORDS
SOURCE
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 21)
AUTHORS de Both, G. and de Beuckeleer, M.
TITLE Hybrid winter oilseed rape and methods for producing same
JOURNAL Patent: WO 0141558-A 12 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES
source
1..21
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer 201"

BASE COUNT      6 a      5 c      4 g      6 t
ORIGIN

Query Match
Best Local Similarity 92.4%; Score 19.4; DB 6; Length 21;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21
|||||
Db 1 GCTGGACTATAACTTGAC 21

RESULT 3
ARGWTUB
LOCUS ARGWTUB 249 bp DNA SYN 02-APR-1988
DEFINITION Soybean beta-1-tubulin gene fused to Ti plasmid unit 7 3UTR.
ACCESSION X05579
VERSION X05579.1 GI:58087
KEYWORDS beta-tubulin; fusion gene; plasmid.
SOURCE
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 249)
AUTHORS Guiltinan, M.J., Velten, J., Bustos, M.M., Cyr, R.J., Schell, J. and Fosket, D.E.
TITLE The expression of a chimeric soybean beta-tubulin gene in tobacco
JOURNAL Mol. Gen. Genet. 207, 328-334 (1987)
FEATURES
source
1..249
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="fusion product (17AA); Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/transl_table=11
/protein_id="CAA29084.1"
/db_xref="GI:4376141"
/translation="MAAAWASSNWSTDPKMS"

CDS
1..51
/note="synthetic construct"

```

---

```

misc_feature 1..13
/note="beta-1-tubulin sequence"
misc_feature 14..28
/note="pUC 13 polylinker"
misc_feature 29..35
/note="Sal I linker"
misc_feature 37
/note="theroretical fusion junction (24) with gene 7 of Ti plasmid"
misc_feature 173..178
/note="put.polyA signal"
polyA_site 198
/note="polyA site"
misc_feature 220..225
/note="put.polyA signal"
BASE COUNT      76 a      47 c      38 g      88 t
ORIGIN

Query Match
Best Local Similarity 92.4%; Score 19.4; DB 12; Length 249;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21
|||||
Db 189 GCTGGACTATAACTTGAC 209

RESULT 4
AXI27755/c
LOCUS AXI27755 415 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 8 from Patent WO0131042.
ACCESSION AXI27755
VERSION AXI27755.1 GI:14134402
KEYWORDS
SOURCE
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 415)
AUTHORS Weston, B. and de Beuckeleer, M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 8 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source
1..415
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="5' border flanking region of elite event MS-B2"
misc_feature 1..234
/note="plant DNA"
misc_feature 235..415
/note="T-DNA"
BASE COUNT      154 a      55 c      70 g      136 t
ORIGIN

Query Match
Best Local Similarity 92.4%; Score 19.4; DB 6; Length 415;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21
|||||
Db 335 GCTGGACTATAACTTGAC 315

RESULT 5
AXI27757
LOCUS AXI27757 416 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 10 from Patent WO0131042.
ACCESSION AXI27757
VERSION AXI27757.1 GI:14134404
KEYWORDS
SOURCE
synthetic construct.

```



ORGANISM synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 416)  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES Location/Qualifiers  
source  
1..416  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="3' border flanking region of elite event MS-B2"  
misc\_feature 1..193  
/note="T-DNA"  
misc\_feature 194..416  
/note="plant DNA"  
BASE COUNT 137 a 72 c 54 g 152 t 1 others  
ORIGIN  
Query Match 92.4%; Score 19.4; DB 6; Length 416;  
Best Local Similarity 95.2%; Pred. NO. 12;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 gcttgactataacttgac 21  
|||||  
Db 81 GCTTGGACTATAATACCTGC 101  
RESULT 6  
AXI72478/c AXI72478 694 bp DNA PAT 03-JUL-2001  
LOCUS AXI72478 Sequence 39 from Patent WO0141558.  
DEFINITION AXI72478  
ACCESSION AXI72478  
VERSION AXI72478.1 GI:14597590  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 694)  
AUTHORS de Both,G. and de Beuckeleer,M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 39 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES Location/Qualifiers  
source  
1..694  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="sequence comprising the 5' flanking region of  
RF-BN1 in WOSR"  
BASE COUNT 298 a 81 c 73 g 242 t  
ORIGIN  
Query Match 92.4%; Score 19.4; DB 6; Length 694;  
Best Local Similarity 95.2%; Pred. NO. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 gcttgactataacttgac 21  
|||||  
Db 694 GCTTGGACTATAATACCTGC 674  
RESULT 7  
ATRN7 ATRN7 831 bp DNA BCT 02-SEP-1999  
LOCUS ATRN7 Agrobacterium tumefaciens gene encoding 'transcript 7' mRNA for a  
DEFINITION protein with unknown function.  
ACCESSION V00090  
VERSION V00090.1 GI:39180  
KEYWORDS unidentified reading frame.  
SOURCE Agrobacterium tumefaciens.  
ORGANISM Agrobacterium tumefaciens

Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group:  
Rhizobiaceae; Rhizobium.  
REFERENCE 1 (bases 1 to 831)  
AUTHORS Dhaese,P., De Greve,H., Gielen,J., Seurinck,J., Van Montagu,M.M.  
and Schell,J.  
TITLE Identification of sequences involved in the polyadenylation of  
higher plant nuclear transcripts using Agrobacterium T-DNA genes as  
models  
JOURNAL EMBO J. 2, 419-426 (1983)  
REFERENCE 2 (bases 76 to 100)  
AUTHORS Dhaese,P.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAY-1983) to the EMBL/GenBank/DBJ databases  
COMMENT Data kindly reviewed (27-MAY-1983) by Dhaese P.  
FEATURES Location/Qualifiers  
source  
1..831  
/organism="Agrobacterium tumefaciens"  
/strain="(octopine TL-DNA)"  
/db\_xref="taxon:358"  
75..81  
/note="CG [1] revised CCAGAGG [2]"  
/citation=[1]  
/citation=[2]  
99..101  
/note="CCA [1] revised CTA [2]"  
/citation=[1]  
/citation=[2]  
132..673  
/note="transcript 7 (alternate)"  
132..672  
/note="transcript 7"  
148..528  
/note="unknown gene (148 is 1st base in codon) (525 is 3rd  
base in codon)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA23429.1"  
/db\_xref="GI:39181"  
/db\_xref="SWISS-PROT:P03867"  
/translation="MNFADTPLASLDLDNACEEFIKTVGASPOLETGVIOTNNGLLY  
LYGKSLSORIHDTHLAFKEKEELSFTTIKPAEMKAQSDTTYVAIFQSNFLCVSN  
PEKGLRCHNRPFLYPIVAHGSM"  
BASE COUNT 262 a 176 c 138 g 255 t  
ORIGIN  
Query Match 92.4%; Score 19.4; DB 1; Length 831;  
Best Local Similarity 95.2%; Pred. NO. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 gcttgactataacttgac 21  
|||||  
Db 664 GCTTGGACTATAATACCTGC 684  
RESULT 8  
LOCUS ATTDNA 878 bp DNA BCT 25-MAR-1996  
DEFINITION Agrobacterium tumefaciens crown gall tumor T-DNA from Ti (tumor  
inducing) plasmid pTiA6.  
ACCESSION X00431  
VERSION X00431.1 GI:39150  
KEYWORDS plasmid.  
SOURCE Agrobacterium tumefaciens.  
ORGANISM Agrobacterium tumefaciens  
Bacteria; Proteobacteria;  
Rhizobiaceae; Rhizobium.  
REFERENCE 1 (bases 1 to 878)  
AUTHORS McPherson,J.C.  
TITLE DNA sequence analysis of crown gall tumor T-DNA encoding the 0.7 kb  
transcript  
JOURNAL Nucleic Acids Res. 12 (5), 2317-2325 (1984)  
MEDLINE 84169535

```
FEATURES
  source          Location/Qualifiers
1..878
  /organism="Agrobacterium tumefaciens"
  /strain="plasmid pRIa6"
  /db_xref="taxon:358"
60..66
  /note="TATA-box"
68..613
  /note="polyadenylation signal"
109..489
  /note="unidentified reading frame"
  /codon_start=1
  /transl_table=11
  /protein_id="CAA25129.1"
  /db_xref="GI:39151"
  /db_xref="SWISS-PROT:P03867"
  /translation="MNFADTFLASLDLDMACEEFIKTYGASPOLETGEVIQTNNGLLY
LYGKGSLSQRHDLFLAFKEKEELSFTTIKPAEMKAQOSDLTYVVAIFQSNFLCVSN
PERGLRCHNRPFLYPIVAGHMS"
659..664
  /note="polyadenylation signal"
288 a 189 c 139 g 262 t
misc_feature
BASE COUNT 288 a 189 c 139 g 262 t
ORIGIN

Query Match          92.4%; Score 19.4; DB 1; Length 878;
Best Local Similarity 95.2%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21
|||||
Db 625 GCTTGGACTATAACTCTGAC 645

RESULT 9
AX172475/c
LOCUS AX172475 909 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 36 from Patent WO0141558.
ACCESSION AX172475
VERSION AX172475.1 GI:14597587
KEYWORDS
SOURCE
  synthetic construct.
  artificial sequence.
REFERENCE 1 (bases 1 to 909)
AUTHORS de Both,G. and de Beuckeleer,M.
TITLE Hybrid winter oilseed rape and methods for producing same
JOURNAL Patent: WO 0141558-A 36 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES
  source          Location/Qualifiers
1..909
  /organism="synthetic construct"
  /db_xref="taxon:32630"
  /note="sequence comprising the 5' flanking region of
MS-BNI in WOSR"
BASE COUNT 309 a 134 c 173 g 293 t
ORIGIN

Query Match          92.4%; Score 19.4; DB 6; Length 909;
Best Local Similarity 95.2%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21
|||||
Db 909 GCTTGGACTATAACTCTGAC 889

RESULT 10
AX10942
LOCUS AX10942 1037 bp DNA PAT 27-SEP-1993
DEFINITION Nucleotide sequence 4 from patent number DE3920034.
ACCESSION AX10942
```

```
VERSION A10942.1 GI:492369
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1037)
AUTHORS
JOURNAL Patent: DE 3920034-A 4 31-MAY-1990;
FEATURES
  source          Location/Qualifiers
1..1037
  /organism="unidentified"
  /db_xref="taxon:32644"
BASE COUNT 338 a 174 c 166 g 359 t
ORIGIN

Query Match          92.4%; Score 19.4; DB 6; Length 1037;
Best Local Similarity 95.2%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21
|||||
Db 921 GCTTGGACTATAACTCTGAC 941

RESULT 11
AX172463/c
LOCUS AX172463 1077 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 24 from Patent WO0141558.
ACCESSION AX172463
VERSION AX172463.1 GI:14597575
KEYWORDS
SOURCE
  synthetic construct.
  artificial sequence.
REFERENCE 1 (bases 1 to 1077)
AUTHORS de Both,G. and de Beuckeleer,M.
TITLE Hybrid winter oilseed rape and methods for producing same
JOURNAL Patent: WO 0141558-A 24 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES
  source          Location/Qualifiers
1..1077
  /organism="synthetic construct"
  /db_xref="taxon:32630"
  /note="sequence comprising the 5' flanking region of
RF-BNI"
misc_feature 1..45
  /note="pGEM -T vector"
misc_feature 1061..1077
  /note="pGEM -T vector"
BASE COUNT 430 a 148 c 151 g 347 t 1 others
ORIGIN

Query Match          92.4%; Score 19.4; DB 6; Length 1077;
Best Local Similarity 95.2%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21
|||||
Db 980 GCTTGGACTATAACTCTGAC 960

RESULT 12
A10939
LOCUS A10939 1085 bp DNA PAT 27-SEP-1993
DEFINITION Nucleotide sequence 1 from patent number DE3920034.
ACCESSION A10939
VERSION A10939.1 GI:492367
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
```

REFERENCE 1 (bases 1 to 1085)  
AUTHORS  
JOURNAL Patent: DE 3920034-A 1 31-MAY-1990;  
FEATURES Location/Qualifiers  
1. .1085  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 369 a 218 c 155 g 343 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1085;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 gcttgactataacttgac 21  
|||||  
Db 969 GCTTGGACTATATACCTGAC 989

RESULT 13  
A10943  
LOCUS A10943 1160 bp DNA PAT 27-SEP-1993  
DEFINITION Nucleotide sequence 5 from patent number DE3920034.  
ACCESSION A10943  
VERSION A10943.1 GI:492370  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1160)  
AUTHORS  
JOURNAL Patent: DE 3920034-A 5 31-MAY-1990;  
FEATURES Location/Qualifiers  
1. .1160  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 367 a 194 c 188 g 411 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1160;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 gcttgactataacttgac 21  
|||||  
Db 1044 GCTTGGACTATATACCTGAC 1064

RESULT 14  
A18051  
LOCUS A18051 1186 bp DNA PAT 26-JUL-1994  
DEFINITION DNA used as a probe for neo gene seq ID No:2.  
ACCESSION A18051  
VERSION A18051.1 GI:583120  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE artificial sequence.  
1 (bases 1 to 1186)  
AUTHORS  
JOURNAL Patent: WO 9209696-A 2 11-JUN-1992;  
FEATURES Location/Qualifiers  
1. .1186  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
1. .8  
/note="sequence derived from tapetum specific promoter of  
Nicotiana tabacum"  
gene 167..790  
CDS 167..790  
/gene="neomycine phosphotransferase gene"

/gene="neomycine phosphotransferase gene"  
/note="protein sequence is in conflict with the conceptual  
translation"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA01373.1"  
/db\_xref="GI:4529900"  
/translation="NEIQDEAARLSWLTATGVCPCAAVLDVVTEAGROWLLIGVEVPGQD  
LLSSHLAPAEKVSIMADMRRLHTLDPATCFDQAKHRIERARTRMEAGLVDDQDLD  
EEHGLAPAEFLKARLMPDGEDLVVTHGDACLPNIWENGRESGFTDCGRLGVADR  
YDIALATRDIAEELGGEWADRELVLVYGAAPDSQRIAFYRLDDEF"  
1055. 1186  
/note="3' regulatory sequence containing the  
polyadenylation site derived from agrobacterium T-DNA gene  
7"  
polyA\_site  
BASE COUNT 244 a 317 c 325 g 300 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1186;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 gcttgactataacttgac 21  
|||||  
Db 1113 GCTTGGACTATATACCTGAC 1133

RESULT 15  
AR095107  
LOCUS AR095107 1186 bp DNA PAT 08-SEP-2000  
DEFINITION Sequence 2 from patent US 6002070.  
ACCESSION AR095107  
VERSION AR095107.1 GI:10022665  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1186)  
AUTHORS D'Halluin,K. and Gobel,E.  
TITLE Process for transforming monocotyledonous plants  
JOURNAL Patent: US 6002070-A 2 14-DEC-1999;  
FEATURES Location/Qualifiers  
1. .1186  
/organism="unknown"  
BASE COUNT 244 a 317 c 325 g 300 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1186;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 gcttgactataacttgac 21  
|||||  
Db 1113 GCTTGGACTATATACCTGAC 1133

Search completed: February 25, 2002, 18:02:56  
Job time: 18584 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:40 ; Search time 716.55 Seconds  
(without alignments)  
25.126 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21

Sequence: 1 gcttgactataataacttgac 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	22	PCR primer B02, to
2	19.4	92.4	21	22	PCR primer for vec
3	19.4	92.4	415	22	Right (5') border
4	19.4	92.4	416	22	Left (3') border f
5	19.4	92.4	694	22	Nucleotide fragmen
6	19.4	92.4	909	22	Nucleotide fragmen
7	19.4	92.4	1037	11	USP-Promoter-casse
8	19.4	92.4	1077	11	Right flanking reg
9	19.4	92.4	1085	11	Legumin-signalpept
10	19.4	92.4	1160	11	USP-signalpeptide
11	19.4	92.4	1186	13	Chimeric neo gene

12	19.4	92.4	1303	17	AAT393337	Plasmid pT898 (Eco
13	19.4	92.4	3153	21	AAZ29122	plasmid DV131 comp
14	19.4	92.4	3201	12	AAQ14529	pp529 Bt ICP codi
15	19.4	92.4	3201	12	AAQ15144	pVE36 Bt ICP codin
16	19.4	92.4	3336	21	AAZ29121	Plasmid DV130 comp
17	19.4	92.4	3694	21	AAZ29124	Plasmid DV133 used
18	19.4	92.4	3877	21	AAZ29123	Plasmid DV132 used
19	19.4	92.4	4832	22	AAH25423	Nucleotide sequenc
20	19.4	92.4	4946	18	AAT59531	T-DNA of plasmid p
21	19.4	92.4	4946	18	AAT59531	Nucleotide sequenc
22	19.4	92.4	5228	22	AAH25432	plasmid pT817delc
23	19.4	92.4	5349	19	AAZ23239	T-DNA of pT824.
24	19.4	92.4	5864	17	AAT39339	Plasmid pT80113 T-
25	19.4	92.4	5864	17	AAT39339	Plasmid pT80113 T-
26	19.4	92.4	5865	22	AAD06990	Chimeric T-DNA of
27	19.4	92.4	5865	22	AAD06990	Chimeric T-DNA of
28	19.4	92.4	6539	21	AAZ91097	E. coli plasmid pT
29	19.4	92.4	6548	17	AAT39336	Plasmid pT8174 use
30	19.4	92.4	6548	18	AAT61394	Plasmid pT8172. C
31	19.4	92.4	6548	21	AAZ91096	E. coli plasmid pT
32	19.4	92.4	7492	22	AAH86441	Plasmid pT8346. U
33	19.4	92.4	7565	14	AAQ42160	Plasmid pT80212 co
34	19.4	92.4	7599	22	AAH25320	Nucleotide sequenc
35	19.4	92.4	7639	14	AAQ42159	Plasmid pJ0884 con
36	19.4	92.4	24593	6	AAH50226	Sequence of opine
37	19.4	92.4	24596	6	AAH50182	Complete nucleotid
38	16.2	77.1	423	21	AAH02009	Human secreted pro
39	16	76.2	22	20	AAZ28158	Transformed Arabid
40	15.8	75.2	81	16	AAH06276	HIV-1 reverse tran
41	15.8	75.2	359	21	AAH06027	Human secreted pro
42	15.8	75.2	831	20	AAZ10777	Trehalose-6-phosph
43	15.8	75.2	909	20	AAZ22113	Human secreted pro
44	15.8	75.2	1433	15	AAQ65397	Elm C10:O-acyl car
45	15.8	75.2	1433	16	AAQ92306	Elm class II thioe

#### ALIGNMENTS

RESULT 1

AA07001 1  
ID AAD07001 standard; DNA; 21 BP.  
AC AAD07001;  
XX  
XX  
XX 06-AUG-2001 (first entry)  
XX  
XX PCR primer B02, to recognise foreign DNA and flanking sequence of MS-B2.  
XX MS-B2 elite event; transgenic Brassica plant; transformation event;  
XX male-sterility gene; PCR primer; ss.  
XX Unidentified.  
XX  
XX WO200131042-A2.  
XX  
XX 03-MAY-2001.  
XX  
XX 26-OCT-2000; 2000WO-EP10680.  
XX  
XX 29-OCT-1999; 99US-0430497.  
XX  
XX (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
XX Weston B, De Beuckeleer M;  
XX  
XX WPI; 2001-300517/31.  
XX  
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
XX harboring specific transformation events, particularly by presence of  
XX male-sterility gene, at specific location in its genome -  
XX  
XX Claim 1; Page 33; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.  
 XX Sequence 21 BP; 6 A; 4 C; 4 G; 7 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0;

Qy 1 gcttggaactataacttgac 21  
 |||||  
 Db 1 gcttggaactataacttgac 21

RESULT 2  
 AAH25431  
 ID AAH25431 standard; DNA; 21 BP.  
 AC AAH25431;  
 XX  
 DT 22-AUG-2001 (first entry)  
 DE PCR primer for vector fragments in transgenic plant MS-BN1.  
 XX  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.  
 OS Synthetic.  
 XX  
 PN WO200141558-A1.  
 XX  
 PD 14-JUN-2001.

XX 06-DEC-2000; 2000WO-EP12872.  
 PF 08-DEC-1999; 99US-0457037.  
 XX  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 PI De Both G, De Beuckeleer M;  
 XX WPI; 2001-381419/40.  
 DR  
 XX

PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
 PT with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome -  
 XX  
 PS Example 4; Page 43; 98pp; English.

XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAH25429-31 were used to amplify fragments of a vector in  
 CC a transgenic plant which carries the TA29-barnase transgene.

XX Sequence 21 BP; 6 A; 5 C; 4 G; 6 T; 0 other;

Query Match 92.4%; Score 19.4; DB 22; Length 21;  
 Best Local Similarity 95.2%; Pred. No. 1.1;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggaactataacttgac 21  
 |||||  
 Db 1 gcttggaactataacttgac 21

RESULT 3  
 AAD06997/c  
 ID AAD06997 standard; DNA; 415 BP.  
 XX  
 AC AAD06997;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Right (5') border flanking region of elite event MS-B2.  
 XX  
 KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; ds.  
 XX  
 OS Chimeric - Agrobacterium sp.  
 OS Chimeric - Brassica sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..234  
 FT /\*tag= a  
 FT /note= "Corresponds to plant DNA"  
 FT 235..415  
 FT /\*tag= b  
 FT /note= "Corresponds to T-DNA"

XX  
 PN WO200131042-A2.  
 XX  
 PD 03-MAY-2001.  
 XX  
 PF 26-OCT-2000; 2000WO-EP10680.  
 XX  
 PR 29-OCT-1999; 99US-0430497.  
 XX  
 PA (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 PI Weston B, De Beuckeleer M;  
 XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harbouring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX  
 PS Claim 11; Page 51; 53pp; English.  
 CC The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is right (5') border flanking region of elite event  
 CC MS-B2.  
 XX Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Claim 11; Page 51; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is right (5') border flanking region of elite event  
 CC MS-B2.

XX Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 92.4%; Score 19.4; DB 22; Length 415;  
 Best Local Similarity 95.2%; Pred. No. 1.3;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggaactataacttgac 21  
 |||||  
 Db 335 GCTTGGACTATAATACCTGAC 315



PF 06-DEC-2000; 2000WO-EP12872.  
 XX  
 PR 08-DEC-1999; 99US-0457037.  
 XX  
 PA (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 XX De Both G, De Beuckeleer M;  
 XX WPI; 2001-381419/40.  
 DR  
 XX Transgenic winter oilseed rape plants suited for producing hybrid seed  
 PT with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome  
 XX  
 XX Example 4; Page 93; 98pp; English.  
 PS  
 XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC The present sequence represents a fragment from a vector comprising the  
 CC TA29-barnase transgene from transgenic plants.  
 XX  
 SQ Sequence 909 BP; 309 A; 134 C; 173 G; 293 T; 0 other;  
 Query Match 92.4%; Score 19.4; DB 22; Length 909;  
 Best Local Similarity 95.2%; Pred. No. 1.3;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcttgagactataacttgac 21  
 Db 909 GCTTGACTATAACTGAC 889  
 RESULT 7  
 AAQ04705  
 ID AAQ04705 standard; DNA; 1037 BP.  
 XX  
 AC AAQ04705;  
 XX  
 DT 12-OCT-1990 (first entry)  
 XX  
 DE USP-Promoter-cassette USP-Pr.T7.1.  
 KW Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.  
 XX  
 PN DE3920034-A.  
 XX  
 PD 31-MAY-1990.  
 XX  
 PF 20-JUN-1989; 89DE-3920034.  
 XX  
 PR 19-SEP-1988; 88DD-0319887.  
 XX  
 XX (PFLA-) VE KOMB PFLANZENZUC.  
 XX  
 PI Bassuner R, Baumelein H, Muntz K, Hai NV, Wobus U;  
 XX WPI; 1990-172459/23.  
 DR  
 XX Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.  
 PT  
 XX Disclosure; ; pp; German.  
 PS  
 XX The unique BglII-Ort (720-725) site is for ligating foreign DNA and the

CC HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 CC cassette in the T1-vector pGA471. Agrobacterium tumefaciens is  
 CC transfected.  
 CC See also AAQ04703-Q04706.  
 XX  
 SQ Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 92.4%; Score 19.4; DB 11; Length 1037;  
 Best Local Similarity 95.2%; Pred. No. 1.4;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcttgagactataacttgac 21  
 Db 921 gcttgagactataacttgac 941  
 RESULT 8  
 AAH25439/C  
 ID AAH25439 standard; DNA; 1077 BP.  
 XX  
 AC AAH25439;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 DE Right flanking region in transgenic plant BN-RF1.  
 KW  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; barstar gene; ss.  
 OS Synthetic.  
 XX  
 PN WO200141558-A1.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 06-DEC-2000; 2000WO-EP12872.  
 XX  
 PR 08-DEC-1999; 99US-0457037.  
 XX  
 PA (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 PI De Both G, De Beuckeleer M;  
 XX  
 DR WPI; 2001-381419/40.  
 XX  
 PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
 PT with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome  
 XX  
 PS Claim 58; Page 88-89; 98pp; English.  
 CC  
 CC The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC The present sequence represents the right flanking region of  
 CC a vector in a transgenic plant which carries the TA29-barstar  
 CC transgene.  
 XX  
 SQ Sequence 1077 BP; 430 A; 148 C; 151 G; 347 T; 1 other;

Query Match 92.4%; Score 19.4; DB 22; Length 1077;  
 Best Local Similarity 95.2%; Pred. No. 1.4;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcttgagactataacttgac 21



Db 980 GCTTGGACTATAATACCTGAC 960  
|||||

## RESULT 9

AAQ04703  
ID AAQ04703 standard; DNA; 1085 BP.

XX AC AAQ04703;

XX DT 12-OCT-1990 (first entry)

XX DE Legumin-signalpeptide cassette Le-Sig.T7.

XX KW Foreign DNA incorporation; recombinant DNA techniques;  
XX KW higher plant genome; legumin; B4 gene; signalpeptide; Le-Sig.T7.; ss.

XX FH Key Location/Qualifiers

XX CDS 747..814

XX FT /\*tag= a  
XX FT /product=Legumin-signalpeptide

XX PN DE3920034-A.

XX PD 31-MAY-1990.

XX PF 20-JUN-1989; 89DE-3920034.

XX PR 19-SEP-1988; 88DD-0319887.

XX PA (PFLA-) VE KOMB PFLANZENZUC.

XX PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

XX DR WPI; 1990-172459/23.

XX DR P-PSDB; AAR05198.

XX PT Incorporation of DNA into higher plant genome - by specified  
XX PT recombinant DNA techniques.

XX PS Disclosure; ; pp; German.

XX CC The Legumin gene B4 is used. The unique BglII-Ort (815-820) site is for  
XX CC ligating foreign DNA and the HindIII-Ort in the 3' polylinker  
XX CC (1080-1085) for cloning the cassette in the Ti-vector pGA471.  
XX CC The cassette is cloned into the binary Ti-vectors pGA471 and  
XX CC Agrobacterium tumefaciens is transfected.  
XX CC See also AAQ04703-Q04706.

XX SQ Sequence 1085 BP; 368 A; 219 C; 155 G; 343 T; 0 other;

## Query Match

Best Local Similarity 92.4%; Score 19.4; DB 11; Length 1085;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21

Db 969 gcttggactataactcgac 989  
|||||

## RESULT 10

AAQ04706

ID AAQ04706 standard; DNA; 1160 BP.

XX AC AAQ04706;

XX DT 12-OCT-1990 (first entry)

XX DE USP-signalpeptide cassette USP-Sig.T7.

XX KW Foreign DNA incorporation; recombinant DNA techniques;  
XX KW higher plant genome; signalpeptide; USP-Sig.T7.; ss.

XX FH Key Location/Qualifiers  
XX CDS 708..877  
XX FT /\*tag= a  
XX FT /product=signalpeptide  
XX FT 747..817

XX PN DE3920034-A.

XX PD 31-MAY-1990.

XX PF 20-JUN-1989; 89DE-3920034.

XX PR 19-SEP-1988; 88DD-0319887.

XX PA (PFLA-) VE KOMB PFLANZENZUC.

XX PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

XX DR WPI; 1990-172459/23.

XX DR P-PSDB; AAR05199.

XX PT Incorporation of DNA into higher plant genome - by specified  
XX PT recombinant DNA techniques.

XX PS Disclosure; ; pp; German.

XX CC The unique BglII-Ort (890-895) site is for  
XX CC ligating foreign DNA and the HindIII-Ort in the 3' polylinker  
XX CC (1155-1160) for cloning the cassette in the Ti-vector pGA471.  
XX CC The cassette is cloned into the binary Ti-vectors pGA471 and  
XX CC Agrobacterium tumefaciens is transfected.  
XX CC See also AAQ04703-Q04706.

XX SQ Sequence 1160 BP; 369 A; 192 C; 188 G; 411 T; 0 other;

## Query Match

Best Local Similarity 92.4%; Score 19.4; DB 11; Length 1160;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21

Db 1044 gcttggactataactcgac 1064  
|||||

## RESULT 11

AAQ25707

ID AAQ25707 standard; DNA; 1186 BP.

XX AC AAQ25707;

XX DT 07-DEC-1992 (first entry)

XX DE Chimeric neo gene probe.

XX KW Nicotiana tabacum; neomycine phosphotransferase; Agrobacterium; ss.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT promoter 1..8

XX FT /\*tag= a  
XX FT /note= "sequence derived from tapetum specific  
promoter of Nicotiana tabacum"

XX CDS 9..790

XX FT /\*tag= b  
XX FT /product= neomycine\_phosphotransferase

XX FT 791..1186

XX FT /\*tag= c

XX FT /note= "3' regulatory sequence contg. the  
polyadenylation site derived from  
Agrobacterium T-DNA gene 7"

```

XX PN W09209696-A.
XX PD 11-JUN-1992.
XX PF 21-NOV-1991; 91WO-EP02198.
XX PR 23-NOV-1990; 90EP-0403332.
XX PR 08-JUL-1991; 91EP-0401888.
XX PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PT Dhalluin K, Goebel E;
XX DR WPI; 1992-217075/26.
XX PT Transforming monocotyledonous plants e.g. cereals - comprises
XX PT wounding and/or degrading cells of intact plant tissue or
XX PS embryogenic callus
XX PS Disclosure; Page 60; 76pp; English.
XX CC Two transformed corn plants were analysed by means of Southern
XX CC hybridisation. As a probe, a 1184 bp EcoRI-HindIII fragment derived
XX CC from another plasmid was used. The sequence of that plasmid is
XX CC given below. Results showed that at least a chimeric neo gene was
XX CC integrated into the plant genomic DNA.
XX SQ Sequence 1186 BP; 244 A; 317 C; 325 G; 300 T; 0 other;

Query Match 92.4%; Score 19.4; DB 13; Length 1186;
Best Local Similarity 95.2%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgacctataactctgac 21
   |||||
Db 1113 gcttgacctataactctgac 1133

RESULT 12
AAT39337
ID AAT39337 standard; DNA; 1303 BP.
AC AAT39337;
XX
XX 22-JAN-1997 (first entry)
XX Plasmid pTS88 (EcoRI-HindIII fragment).
XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
XX transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FH misc_feature 1..35
XX FT /*tag= a
XX FT /label= pGEM2
XX FT /notes= "polylinker of pGEM2"
XX FT 36..694
XX FT /*tag= b
XX FT /label= p35S
XX FT /function= 35S promoter of cauliflower mosaic virus
XX CDS strain CM1841
XX FT 695..967
XX FT /*tag= c
XX FT /label= barstar
XX FT /product= Bacillus amyloliquefaciens barstar
XX FT 968..1287
XX FT /*tag= d
XX FT /label= 3'g7
XX FT /function= region containing polyadenylation signal

```

```

FT misc_feature 1288..1303 of gene 7 og Agrobacterium T-DNA
FT /*tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX
XX W09626283-A1.
XX PD 29-AUG-1996.
XX XX
XX PF 21-FEB-1996; 96WO-EP00722.
XX PR 21-FEB-1995; 95EP-0400364.
XX PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PT Botterman J, Cornelissen M, Michiels F;
XX DR WPI; 1996-402373/40.
XX XX
XX PT Prodn. of male sterile plants by transforming with a chimaeric
XX PT construct - comprising a male sterility DNA e.g. barnase and a
XX PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX PT for generating hybrid cultivars
XX PS Example 1; Page 38; 56pp; English.
XX CC
XX CC The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains
XX CC barstar DNA under control of a 35S promoter. The plasmid was
XX CC used with pTS174 (see also AAT39336) contg. barnase DNA under
XX CC control of the stamen-specific promoter E1 to produce male sterile
XX CC rice cv. Kochinibiki transgenic plants, and with plasmid pV136
XX CC (see also AAT39338) contg. barnase DNA under control of the stamen-
XX CC specific pCA55 promoter to produce male sterile maize plants.
XX CC Expression of barnase (a ribonuclease) in the stamen leads to male
XX CC sterility. Constitutive expression of barstar counteracts possible
XX CC low level expression of barnase DNA in non-stamen tissue.
XX SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 92.4%; Score 19.4; DB 17; Length 1303;
Best Local Similarity 95.2%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgacctataactctgac 21
   |||||
Db 1178 gcttgacctataactctgac 1198

RESULT 13
AAZ29122
ID AAZ29122 standard; DNA; 3153 BP.
XX
XX AC AAZ29122;
XX XX
XX DT 21-FEB-2000 (first entry)
XX DE
XX DE Plasmid DV131 comprising L3/EGFP:NPTII/Tr7 expression cassette.
XX KW Transgenic seed; marker; aleurone-specific promoter; plasmid DV131;
XX KW GFP:NPTII fusion protein construct; L3/EGFP:NPTII/Tr7; maize R gene;
XX KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;
XX KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene; assay;
XX KW Green fluorescent protein; GFP; gene fusion; selection; screening;
XX KW expression; automated seed screening technique; screenable marker;
XX KW transformant; embryogenic tissue; implementation; ds.
XX OS Synthetic.
XX XX
XX PN W0960129-A1.
XX PD 25-NOV-1999.

```

XX 18-MAY-1999; 99WO-US11023.  
 XX 18-MAY-1998; 98US-0080625.  
 XX (DEKA-) DEKALB GENETICS CORP.  
 XX Kriz AL, Spencer TM;  
 XX WPI; 2000-072441/06.  
 XX Screenable marker genes useful for identification of transgenic seeds  
 XX for plant breeding -  
 XX  
 XX Example 1; Page 164-166; 182pp; English.  
 XX The present DNA sequence is the plasmid DVI131, that is used in the  
 CC generation of GFP:NPTII fusion protein constructs. It contains an  
 CC expression cassette comprising, a promoter from the maize L3 oleosin  
 CC gene, the coding sequence of EGFP:NPTII translational fusion, excised  
 CC from DVI126 and the Tr7 terminator. This plasmid is used to carry a  
 CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to the gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.  
 XX  
 XX Sequence 3153 BP; 801 A; 857 C; 756 G; 739 T; 0 other;

Query Match 92.4%; Score 19.4; DB 21; Length 3153;  
 Best Local Similarity 95.2%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataataacttgac 21  
 |||||  
 Db 2759 gcttggactataataacttgac 2779

RESULT 14  
 AAQ14529  
 ID AAQ14529 standard; DNA; 3201 BP.

XX AC AAQ14529;

XX DT 27-JAN-1992 (first entry)

XX DE pPS029 Bt ICP coding sequence.

XX KW Bacillus thuringiensis; insecticidal crystal protein; ICP;  
 XX deletion; ss.

XX OS Synthetic.

XX PN WO9116432-A.

XX PD 31-OCT-1991.

XX PF 17-APR-1991; 91WO-EP00733.

XX PR 18-APR-1990; 90EP-0401055.

XX PA (PLAN-) PLANT GENETIC SYST.

XX PI Cornelissen M, Soetaert P, Stam M, Dockx J;

XX DR WPI; 1991-339820/46.

XX

PT Modified Bacillus thuringiensis insecticidal crystal protein  
 PT genes - having A and T sequences changed to G and C sequences  
 XX encoding same amino acids, for increased expression levels  
 XX PS Disclosure; Fig 6(c); 78pp; English.  
 XX "n" in the sequence refers to not known nucleotides.  
 CC pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-  
 CC terminal modification and the internal modification of the Bt ICP  
 CC coding sequence.  
 CC See also AAQ14529, AAQ15142-44.  
 XX  
 XX SQ Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;

Query Match 92.4%; Score 19.4; DB 12; Length 3201;  
 Best Local Similarity 95.2%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataataacttgac 21  
 |||||  
 Db 3032 gcttggactataataacttgac 3052

RESULT 15

AAQ15144

ID AAQ15144 standard; DNA; 3201 BP.

XX AC AAQ15144;

XX DT 27-JAN-1992 (first entry)

XX DE pVE36 Bt ICP coding sequence.

XX KW Bacillus thuringiensis; insecticidal crystal protein; ICP;  
 XX deletion; ss.

XX OS Synthetic.

XX PN WO9116432-A.

XX PD 31-OCT-1991.

XX PF 17-APR-1991; 91WO-EP00733.

XX PR 18-APR-1990; 90EP-0401055.

XX PA (PLAN-) PLANT GENETIC SYST.

XX PI Cornelissen M, Soetaert P, Stam M, Dockx J;

XX DR WPI; 1991-339820/46.

XX PT Modified Bacillus thuringiensis insecticidal crystal protein  
 XX genes - having A and T sequences changed to G and C sequences  
 XX encoding same amino acids, for increased expression levels

XX PS Disclosure; Fig 6(c); 78pp; English.

XX "n" in the sequence refers to not known nucleotides.

CC pPS029 (AAQ14529) is identical to pVE36, but carries both the amino-  
 CC terminal modification and the internal modification of the Bt ICP  
 CC coding sequence.

CC See also AAQ14529, AAQ15142-44.

XX SQ Sequence 3201 BP; 915 A; 651 C; 700 G; 930 T; 5 other;

Query Match 92.4%; Score 19.4; DB 12; Length 3201;  
 Best Local Similarity 95.2%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataataacttgac 21

Tue Feb 26 09:11:31 2002

us-09-698-903b-12.rng

Page 8

Db 3105 gcttggaactatatacctgac 3125  
|||||

Search completed: February 25, 2002, 18:17:41  
Job time: 16699 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:51 ; Search time 301.6 Seconds  
(without alignments)  
15.769 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21

Sequence: 1 gcttgactataacttgac 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/1na/PTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/1na/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	92.4	1186	1	US-08-064-121-2
2	19.4	92.4	1186	1	US-08-478-015-2
3	19.4	92.4	1186	3	US-08-475-975-2
4	19.4	92.4	1186	3	US-09-084-889-2
5	19.4	92.4	1303	3	US-08-894-440-2
6	19.4	92.4	3153	4	US-09-080-625-3
7	19.4	92.4	3200	1	US-08-453-104-23
8	19.4	92.4	3200	2	US-08-694-824-23
9	19.4	92.4	3201	1	US-08-453-104-22
10	19.4	92.4	3201	2	US-08-694-824-22
11	19.4	92.4	3336	4	US-09-080-625-2
12	19.4	92.4	3694	4	US-09-080-625-5
13	19.4	92.4	3877	4	US-09-080-625-4
14	19.4	92.4	4946	3	US-08-817-188-1
15	19.4	92.4	5560	3	US-08-817-188-5
16	19.4	92.4	5864	3	US-08-894-440-4
17	19.4	92.4	5864	3	US-08-894-440-4
18	19.4	92.4	6548	3	US-08-894-440-1
19	19.4	92.4	6548	3	US-08-817-188-2
20	19.4	92.4	7566	2	US-08-232-016-23
21	19.4	92.4	7639	2	US-08-232-016-22
22	19.4	92.4	7811	2	US-08-549-680A-5
23	19.4	92.4	24595	6	5428147-1
24	15.8	75.2	81	1	US-08-238-863-88
25	15.8	75.2	81	1	US-08-443-407-88
26	15.8	75.2	81	5	PCT-US95-05600-232
27	15.8	75.2	1433	1	US-07-968-971A-11

c 28	15.8	75.2	1433	1	US-08-383-756-5	Sequence 5, Appl
c 29	15.8	75.2	1433	1	US-08-424-406-2	Sequence 2, Appl
c 30	15.8	75.2	1433	1	US-08-464-523B-8	Sequence 8, Appl
c 31	15.8	75.2	1433	2	US-08-460-898-5	Sequence 5, Appl
c 32	15.2	72.4	262	4	US-09-437-457-13	Sequence 13, Appl
c 33	15.2	72.4	1618	1	US-08-410-540-1	Sequence 1, Appl
c 34	15.2	72.4	1641	4	US-08-659-254-19	Sequence 19, Appl
c 35	15.2	72.4	2811	4	US-08-482-073-3	Sequence 3, Appl
c 36	15.2	72.4	2813	2	US-08-344-155C-99	Sequence 99, Appl
c 37	15.2	72.4	2813	4	US-09-009-490A-90	Sequence 90, Appl
c 38	15.2	72.4	3080	4	US-08-482-073-4	Sequence 4, Appl
c 39	15.2	72.4	4016	1	US-08-410-540-3	Sequence 3, Appl
c 40	14.8	70.5	1925	4	US-08-894-324A-1	Sequence 1, Appl
c 41	14.8	70.5	2778	1	US-08-202-054-1	Sequence 1, Appl
c 42	14.8	70.5	2778	1	US-08-446-923-1	Sequence 1, Appl
c 43	14.6	69.5	3390	4	US-09-550-338-1	Sequence 1, Appl
c 44	14.4	68.6	1498	1	US-07-965-688A-1	Sequence 1, Appl
c 45	14.4	68.6	1498	2	US-08-950-433-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-064-121-2  
; Sequence 2, Application US/08064121  
; Patent No. 5641664  
; GENERAL INFORMATION:  
; APPLICANT: D'HALLUIN, Kathleen  
; APPLICANT: GOBEL, Elke  
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/064,121  
; FILING DATE: 24-MAY-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 90403332.1  
; FILING DATE: 23-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91401888.2  
; FILING DATE: 08-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crane-Feuiv, Sharon E  
; REGISTRATION NUMBER: 36,113  
; REFERENCE/DOCKET NUMBER: 010830-043  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1186 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORGANISM: probe

```

; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPPII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
US-08-064-121-2

```

```

Query Match          92.4%; Score 19.4; DB 1; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 gcttgactataactctgac 21
|||||
Db 1113 GCTTGACTATAACTCTGAC 1133

```

```

RESULT 2
US-08-478-015-2
; Sequence 2, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P. O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPPII
; OTHER INFORMATION: /note= "coding sequence of neomycine phosphotransferase ge
US-08-478-015-2

```

```

Query Match          92.4%; Score 19.4; DB 1; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 gcttgactataactctgac 21
|||||
Db 1113 GCTTGACTATAACTCTGAC 1133

```

```

RESULT 3
US-08-475-975-2
; Sequence 2, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:

```

```
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-08-475-975-2

Query Match 92.4%; Score 19.4; DB 3; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataactctgac 21
|||||
Db 1113 GCTTGGACTATAATACCTGAC 1133

RESULT 4
US-09-084-889-2
; Sequence 2, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; FILING DATE:
; FILING DATE: 08/064,121
; PRIOR APPLICATION DATA: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-09-084-889-2

Query Match 92.4%; Score 19.4; DB 3; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataactctgac 21
|||||
Db 1113 GCTTGGACTATAATACCTGAC 1133

RESULT 5
US-08-894-440-2
; Sequence 2, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2:
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
; OTHER INFORMATION: fragment of pTS88
```

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(35)
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36)..(694)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain
; OTHER INFORMATION: CM1841 (P35S)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (695)..(967)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amy1oliquefaciens
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (968)..(1287)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1288)..(1303)
; OTHER INFORMATION: polylinker of pGEM2
;
; US-08-894-440-2
;
; Query Match          92.4%; Score 19.4; DB 3; Length 1303;
; Best Local Similarity 95.2%; Pred. No. 0.3;
; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 gctggactataactgtac 21
;      |||||
; Db 1178 gctggactataactgtac 1198
;
; RESULT 6
; US-09-080-625-3
; Sequence 3, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; APPLICANT: Spencer, T. Michael
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/080,625
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
;
; Query Match          92.4%; Score 19.4; DB 3; Length 1303;
; Best Local Similarity 95.2%; Pred. No. 0.3;
; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 gctggactataactgtac 21
;      |||||
; Db 1178 gctggactataactgtac 1198
;
; RESULT 7
; US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,104
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
;
; US-08-453-104-23
;
; Query Match          92.4%; Score 19.4; DB 1; Length 3200;
; Best Local Similarity 95.2%; Pred. No. 0.32;
; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
;
; TOPOLOGY: linear
; US-09-080-625-3
;
; Query Match          92.4%; Score 19.4; DB 4; Length 3153;
; Best Local Similarity 95.2%; Pred. No. 0.32;
; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 gctggactataactgtac 21
;      |||||
; Db 2759 GCTTGGACTATAACTGTAC 2779
;
; RESULT 7
; US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,104
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
;
; US-08-453-104-23
;
; Query Match          92.4%; Score 19.4; DB 1; Length 3200;
; Best Local Similarity 95.2%; Pred. No. 0.32;
; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```



Qy 1 gcttgactataacttgac 21  
|||||  
Db 3032 GCTTGACTATAACTCGAC 3052

## RESULT 8

US-08-694-824-23  
; Sequence 23, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 92.4%; Score 19.4; DB 2; Length 3200;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21  
|||||  
Db 3032 GCTTGACTATAACTCGAC 3052

## RESULT 9

US-08-453-104-22  
; Sequence 22, Application US/08453104

; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 92.4%; Score 19.4; DB 1; Length 3201;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21  
|||||  
Db 3105 GCTTGACTATAACTCGAC 3125

## RESULT 10

US-08-694-824-22  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

;; TITLE OF INVENTION: IN PLANT CELLS  
;; NUMBER OF SEQUENCES: 23  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: George Mason Bldg., Washington & Prince Sts.  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/694,824  
;; FILING DATE: 09-AUG-1996  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/937,869  
;; FILING DATE: 16-DEC-1992  
;; APPLICATION NUMBER: GB 90401055.0  
;; FILING DATE: 18-APR-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Rea, Teresa S  
;; REGISTRATION NUMBER: 30,427  
;; REFERENCE/DOCKET NUMBER: 010830-032  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 22:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3201 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 2151..2155  
;; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
;; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-22

Query Match 92.4%; Score 19.4; DB 2; Length 3201;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21  
|||||  
Db 3105 GCTTGGACTATAACTGTGAC 3125

RESULT 11  
US-09-080-625-2  
; Sequence 2, Application US/09080625  
; Patent No. 6307123  
; GENERAL INFORMATION:  
; APPLICANT: Kriz, Alan L.  
; APPLICANT: Spencer, T. Michael  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
; TITLE OF INVENTION: IDENTIFICATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/080,625  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hanson, Robert E.  
;; REGISTRATION NUMBER: P-42,628  
;; REFERENCE/DOCKET NUMBER: DEKM:161  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (512) 418-3000  
;; TELEFAX: (512) 474-7577  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3336 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
US-09-080-625-2

Query Match 92.4%; Score 19.4; DB 4; Length 3336;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21  
|||||  
Db 2942 GCTTGGACTATAACTGTGAC 2962

RESULT 12  
US-09-080-625-5  
; Sequence 5, Application US/09080625  
; Patent No. 6307123  
; GENERAL INFORMATION:  
; APPLICANT: Kriz, Alan L.  
; APPLICANT: Spencer, T. Michael  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
; TITLE OF INVENTION: IDENTIFICATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/080,625  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, Robert E.  
; REGISTRATION NUMBER: P-42,628  
; REFERENCE/DOCKET NUMBER: DEKM:161  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3694 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-080-625-5

Query Match 92.4%; Score 19.4; DB 4; Length 3694;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataaacttgac 21  
|||||

Db 3300 GCTTGGACTATAACTCTGAC 3320

## RESULT 13

US-09-080-625-4

; Sequence 4, Application US/09080625

; Patent No. 6307123

; GENERAL INFORMATION:

; APPLICANT: Kriz, Alan L.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE

; TITLE OF INVENTION: IDENTIFICATION

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: TX

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/080.625

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Hanson, Robert E.

; REGISTRATION NUMBER: P-42,628

; REFERENCE/DOCKET NUMBER: DEKM:161

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 418-3000

; TELEFAX: (512) 474-7577

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3877 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-09-080-625-4

Query Match 92.4%; Score 19.4; DB 4; Length 3877;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataaacttgac 21  
|||||

Db 3483 GCTTGGACTATAACTCTGAC 3503

## RESULT 14

US-08-817-188-1/c

; Sequence 1, Application US/08817188

; Patent No. 6074876

; GENERAL INFORMATION:

; APPLICANT: DE BLOCK, MARC

; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR

; FILE REFERENCE: 2121-0127P

; CURRENT APPLICATION NUMBER: US/08/817,188

; CURRENT FILING DATE: 1997-05-15

; EARLIER APPLICATION NUMBER: PCT/EP96/03366

; EARLIER FILING DATE: 1996-07-31

; EARLIER APPLICATION NUMBER: EP 95401844.6  
; EARLIER FILING DATE: 1995-08-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4946

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of

; OTHER INFORMATION: plasmid pTHW107

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((1)..(25))

; OTHER INFORMATION: T-DNA right border (RB)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((97)..(330))

; OTHER INFORMATION: 3'g7; 3' untranslated region containing the

; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium

; OTHER INFORMATION: T-DNA

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((331)..(882))

; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl

; OTHER INFORMATION: transferase

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((883)..(2608))

; OTHER INFORMATION: promoter region of Rubisco small subunit gene of

; OTHER INFORMATION: Arabidopsis thaliana (PSSU)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((2658)..(3031))

; OTHER INFORMATION: 3' nos; 3' untranslated region containing the

; OTHER INFORMATION: polyadenylation signal of the nopaline synthase

; OTHER INFORMATION: gene of Agrobacterium T-DNA

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((3032)..(3367))

; OTHER INFORMATION: barnase: region coding for barnase

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((3368)..(4876))

; OTHER INFORMATION: PTA29; promoter region of TA29 gene of Nicotiana

; OTHER INFORMATION: tabacum

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((4922)..(4946))

; OTHER INFORMATION: LB: T-DNA left border

; US-08-817-188-1

Query Match 92.4%; Score 19.4; DB 3; Length 4946;  
Best Local Similarity 95.2%; Pred. No. 0.34;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataaacttgac 21  
|||||

Db 163 GCTTGGACTATAACTCTGAC 143

## RESULT 15

US-08-817-188-5/c

; Sequence 5, Application US/08817188

; Patent No. 6074876

; GENERAL INFORMATION:

; APPLICANT: DE BLOCK, MARC

; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR

; FILE REFERENCE: 2121-0127P

; CURRENT APPLICATION NUMBER: US/08/817,188

; CURRENT FILING DATE: 1997-05-15

; EARLIER APPLICATION NUMBER: PCT/EP96/03366

```
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CaMV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-LS1 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5067)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CaMV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
```

US-08-817-188-5

```
Query Match          92.4%; Score 19.4; DB 3; Length 5560;
Best Local Similarity 95.2%; Pred. No. 0.34;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggaactataatacttgac 21
   |||||
Db 150 GCITGGACTATATACCTGAC 130
```

Search completed: February 25, 2002, 18:05:52  
Job time: 18600 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:17 ; Search time 8261.74 Seconds  
(without alignments)  
27.314 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21

Sequence: 1 gcttgactataataacttgac 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_estl:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.4	82.9	344	13 AQ117151	AQ117151 HS_2188_A
2	17.4	82.9	633	13 AZ464390	AZ464390 LM0273D15
3	16.8	80.0	291	10 AV035892	AV035892 AV035892
4	16.8	80.0	449	13 AQ172470	AQ172470 HS_3190_B
C 5	16.8	80.0	415	10 AI675738	AI675738 wc40h04.x
6	16.8	80.0	449	11 BE802728	BE802728 sr44g04.y
C 7	16.8	80.0	470	13 AZ133172	AZ133172 OSJNB010
8	16.8	80.0	483	10 AA310103	AA310103 EST180941
9	16.8	80.0	487	13 B74437	B74437 CIT-HSP-202
C 10	16.8	80.0	505	13 AQ666336	AQ666336 HS_5374_A
11	16.8	80.0	537	13 AQ836967	AQ836967 HS_4553_A
12	16.8	80.0	600	11 BE846195	BE846195 232148 BA

C 13	16.8	80.0	854	13 BH132866	BH132866 ENTNI55TF
C 14	16.8	80.0	968	13 AZ670679	AZ670679 ENTH034TF
C 15	16.4	78.1	301	10 BB103479	BB103479 BB103479
16	16.4	78.1	302	10 AA407887	AA407887 EST02498
C 17	16.4	78.1	305	10 BB117588	BB117588 BB117588
18	16.4	78.1	309	10 BB235252	BB235252 BB235252
19	16.4	78.1	310	10 AU021901	AU021901 AU021901
20	16.4	78.1	353	10 AU021997	AU021997 AU021997
C 21	16.4	78.1	376	10 AA823300	AA823300 vp37bl1.r
C 22	16.4	78.1	405	10 AA606880	AA606880 vm87a06.r
23	16.4	78.1	408	10 AU021898	AU021898 AU021898
24	16.4	78.1	432	10 AA763279	AA763279 vvr9d09.r
25	16.4	78.1	453	13 AQ236085	AQ236085 HS_2052_A
26	16.4	78.1	470	10 AU017121	AU017121 AU017121
C 27	16.4	78.1	493	10 AA690887	AA690887 vt32f04.r
28	16.4	78.1	503	10 AU042554	AU042554 AU042554
29	16.4	78.1	514	10 AI315372	AI315372 uj33g12.x
C 30	16.4	78.1	576	10 BE325283	BE325283 NF120D06S
31	16.4	78.1	605	10 AW109828	AW109828 MT329 mou
C 32	16.4	78.1	613	13 AZ411070	AZ411070 LM0184B01
C 33	16.4	78.1	667	13 AQ198178	AQ198178 RPI111-60
C 34	16.4	78.1	678	13 AQ201499	AQ201499 RPI111-60
35	16.4	78.1	863	11 BF978014	BF978014 602147907
36	16.2	77.1	280	10 BB337839	BB337839 BB337839
C 37	16.2	77.1	305	10 AA013302	AA013302 ze28d06.r
C 38	16.2	77.1	411	10 AA554565	AA554565 nk30a02.s
C 39	16.2	77.1	412	11 BF841371	BF841371 RC2-HPI107
C 40	16.2	77.1	423	13 AZ273622	AZ273622 RPI1-23-8
41	16.2	77.1	441	11 H04932	H04932 y174h05.r1
42	16.2	77.1	445	13 AQ170257	AQ170257 HS_5220_A
43	16.2	77.1	458	13 AQ493489	AQ493489 HS_5056_B
44	16.2	77.1	468	10 BE246449	BE246449 TCBAP1E46
45	16.2	77.1	485	13 AG023449	AG023449 Oryza sat

#### ALIGNMENTS

RESULT_1	1				
AQ117151/c					
LOCUS	AQ117151	344 bp	DNA	GSS	22-SEP-1998
DEFINITION	HS_2188_A1_D04_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2188 Col=7 Row=G, DNA sequence.				
ACCESSION	AQ117151				
VERSION	AQ117151.1	GI:3494942			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A.	96 (17),	9739-9744	(1999)	
MEDLINE	99380589				
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 2188 row: G column: 7 Class: BAC ends High quality sequence stop: 344. Location/Qualifiers 1. 344 /organism="Homo sapiens" /db_xref="taxon:9606"				
FEATURES					
source					

```

/clone="Plate-2188 Col-7 Row-G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAcl1; BAC Clones in
E-Coli DH10B"
BASE COUNT 105 a 53 c 85 g 101 t
ORIGIN

```

```

Query Match 82.9%; Score 17.4; DB 13; Length 344;
Best Local Similarity 94.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cttggactataataacttga 20
| | | | | | | | | | | | | | | | | |
Db 167 CATGGACTATAACTTGA 149

```

```

RESULT 2
AZ464390 633 bp DNA GSS 04-OCT-2000
LOCUS
DEFINITION
Clone UUGC1M0273D15 R, DNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
house mouse.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 633)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0273 row: D column: 15
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 633.
Location/Qualifiers
1. .633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0273D15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number

```

## FEATURES

```

source
1. .633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0273D15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number

```

```

Inducible derivative of plasmid pL1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 183 a 133 c 136 g 181 t
ORIGIN

```

```

Query Match 82.9%; Score 17.4; DB 13; Length 633;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cttggactataataacttga 20
| | | | | | | | | | | | | | | | | |
Db 473 CTTGGACTATAAATCTTGA 491

```

```

RESULT 3
AV035892 291 bp mRNA EST 22-NOV-1999
LOCUS
DEFINITION
AV035892 Mus musculus adult C57BL/6J placenta Mus musculus CDNA
clone 1600016H12, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
house mouse.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 291)
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Alzawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
,A., Hayatsu,N., Horii,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H., Owa,C.,
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Soabe,Y., Sugahara
,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tominaga,N.,
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermolabile and its application for the synthesis of full length cDNA
thermostabilization and thermoactivation of thermolabile enzymes by
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. .291
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1600016H12"
/clone_lib="Mus musculus adult C57BL/6J placenta"
/sex="female"
/tissue_type="placenta"
/dev_stage="adult"
BASE COUNT 68 a 56 c 66 g 101 t
ORIGIN

```

```

TITLE
JOURNAL
COMMENT
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermolabile and its application for the synthesis of full length cDNA
thermostabilization and thermoactivation of thermolabile enzymes by
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. .291
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1600016H12"
/clone_lib="Mus musculus adult C57BL/6J placenta"
/sex="female"
/tissue_type="placenta"
/dev_stage="adult"
BASE COUNT 68 a 56 c 66 g 101 t
ORIGIN

```

## FEATURES

```

source
1. .291
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1600016H12"
/clone_lib="Mus musculus adult C57BL/6J placenta"
/sex="female"
/tissue_type="placenta"
/dev_stage="adult"
BASE COUNT 68 a 56 c 66 g 101 t
ORIGIN
Query Match 80.0%; Score 16.8; DB 10; Length 291;
Best Local Similarity 90.0%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 2 cttggactataacttgac 21
    ||||| ||||| ||||| |||||
Db 13 CTTGGACTATATATCGGAC 32

RESULT 4
AQL172470
LOCUS HS_3190_B2_H07_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3190 Col-14 Row-P, DNA sequence. 17-OCT-1998
ACCESSION AQL172470
VERSION AQL172470.1 GI:3569837
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3190 row: P column: 14
Class: BAC ends
High quality sequence stop: 415.
FEATURES
    source
    location/Qualifiers
    1..415
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="CIT Approved Human Genomic Sperm Library D"
    /sex="male"
    /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 77 a 117 c 69 g 151 t 1 others
ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 415;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 cttggactataacttgac 21
    ||||| ||||| ||||| |||||
Db 27 CTTGCAATATATGCTTGAC 46

RESULT 5
A1675738/c
LOCUS A1675738 449 bp mRNA EST 17-DEC-1999
DEFINITION wc40h04.x1 NCI-CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321143 3' similar to gb:W73255_rnal VASCULAR CELL ADHESION PROTEIN 1
PRECUSOR (HUMAN);, mRNA sequence.
ACCESSION A1675738
VERSION A1675738.1 GI:4876218
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 449)

AUTHORS
TITLE
JOURNAL
COMMENT

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
cDNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert length: 567 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 411.
FEATURES
    source
    location/Qualifiers
    1..449
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2321143"
    /clone_lib="NCI-CGAP_Pr28"
    /sex="male"
    /dev_stage="adult"
    /lab_host="DH10B"
    /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones IDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 178 a 77 c 63 g 131 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 10; Length 449;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gcttgactataacttgac 20
    ||||| ||||| ||||| |||||
Db 419 GCTTGGACTATATATTTTA 400

RESULT 6
BE802728/c
LOCUS BE802728 449 bp mRNA EST 20-SEP-2000
DEFINITION sr44904.y1 Gm-c1051 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1051-943 5', mRNA sequence.
ACCESSION BE802728
VERSION BE802728.1 GI:10233840
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 449)
AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project

```

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 High quality sequence stop: 402.

## FEATURES

## source

1. .449  
 Location/Qualifiers  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl051-943"  
 /clone\_1lb="Gm-cl051"  
 /tissue\_type="floral meristematic mRNA"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:  
 XhoI; The cDNA library was constructed from floral  
 meristematic mRNA provided by Dr. Halina Knap of Clemson  
 University. Complementary DNA was synthesized from mRNA  
 using a primer consisting of a poly(dT) sequence with a  
 blunt-ended cDNA fragments followed by XhoI digestion. The  
 cDNA fragments were directionally cloned into the  
 EcoRI-XhoI restriction site of the pBluescript vector. The  
 ligated cDNA fragments were transformed into DH10B host  
 cells (GibcoBRL). This library was constructed in the  
 laboratory of Dr. Randy Shoemaker."  
 BASE COUNT 149 a 73 c 80 g 147 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 11; Length 449;  
 Best Local Similarity 90.0%; Pred. No. 4.8e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 cttggactataactctgac 21  
 |||||  
 Db 358 CTTGGACTATAAAGTTGAC 339

## RESULT 7

AZ133172  
 LOCUS AZ133172 470 bp DNA GSS 02-JUN-2000  
 DEFINITION OSJNB0108P05r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic  
 clone OSJNB0108P05r, DNA sequence.  
 ACCESSION AZ133172  
 VERSION AZ133172.1 GI:8212071  
 KEYWORDS GSS.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 470)

## REFERENCE

## AUTHORS

Wing, R.A. and Dean, R.A.  
 A BAC End Sequencing Framework to Sequence the Rice Genome  
 Unpublished (1998)  
 Contact: Wing RA

## JOURNAL

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GGAAACAGCTATGACCATG

Class: BAC ends

High quality sequence start: 101

High quality sequence stop: 368.

Location/Qualifiers

## FEATURES

## source

1. .470  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="OSJNB0108P05r"  
 /clone\_1lb="CUGI Rice BAC Library (EcoRI)"  
 /tissue\_type="leaf"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pBACIndigo; Site\_1: EcoRI; Site\_2: EcoRI;  
 Rice is the most important food crop in the world. Half of  
 the world population, especially those inhabiting highly  
 populated areas of the humid tropics and subtropics, rely  
 on rice as their primary source of carbohydrate.  
 Monocytiledonous rice is a diploid plant (2n=24) with a  
 haploid genome equivalent of 431 Mbp (Arumuganathan and  
 Earle, 1991). The relatively small genome of rice, three  
 times larger than that of Arabidopsis, makes it suitable  
 for genomic studies. In order to facilitate positional  
 cloning, physical mapping and genome sequencing of rice,  
 we have constructed a BAC library from Oryza sativa,  
 Nipponbare variety using EcoRI as the cloning enzyme. The  
 library contains 55,296 clones with an average insert size  
 of 121 Kb providing approximately 15 haploid genome  
 equivalents. The deep coverage allows the isolation a  
 particular sequence with a probability of 99.9 %. Three  
 high density filters, each containing 18,432 clones  
 (doubly spotted), represent the whole library for colony  
 screening and can be requested from the Clemson University  
 BAC/EST Resource Center (www.genome.clemson.edu)."  
 BASE COUNT 163 a 75 c 94 g 131 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 470;  
 Best Local Similarity 90.0%; Pred. No. 4.9e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcttgactataacttga 20  
 |||||

Db 277 GCTTGGGCTATAACTTGA 296

## RESULT 8

AZ1310103  
 LOCUS AA310103 483 bp mRNA EST 19-APR-1997  
 DEFINITION EST180941 Jurkat T-cells V Homo sapiens cDNA 5' end similar to  
 Wiskott-Aldrich syndrome protein, mRNA sequence.  
 ACCESSION AA310103  
 VERSION AA310103.1 GI:1962431  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

## REFERENCE

## AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 483)  
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult  
 , C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White  
 , O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,  
 Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald  
 , L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,  
 Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,  
 Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
 Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,  
 Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,  
 Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,  
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
 Dimker, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,  
 Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,  
 Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Lisen, H., Raymond, L.,  
 Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon  
 , M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and  
 Venter, J.C.



**TITLE** Initial assessment of human gene diversity and expression patterns  
**JOURNAL** based upon 83 million nucleotides of cDNA sequence  
**MEDLINE** Nature 377 (6547 Suppl), 3-174 (1995)  
**COMMENT** 96026280  
 Other ESTs: THC122434  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

# **FEATURES**

## source

1..483  
 Location/Qualifiers  
 .organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):156165"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Jurkat T-cells V"  
 /cell\_type="n-lymphocyte"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 xhoi"

**BASE COUNT** 113 a 142 c 136 g 85 t 7 others

## ORIGIN

**Query Match** 80.0%; Score 16.8; DB 10; Length 483;  
**Best Local Similarity** 85.7%; Pred. No. 4.9e+02;  
**Matches** 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**Qy** 1 gcttgactataatacttgac 21

**Db** 123 GCTTGGACGAATNCTTGAC 143

## RESULT

**B74437** 9  
**LOCUS** B74437 487 bp DNA GSS 24-OCT-1998  
**DEFINITION** CIT-HSP-2028F18.TF CIT-HSP Homo sapiens genomic clone 2028F18, DNA  
 sequence.

**ACCESSION** B74437

**VERSION** B74437.1 GI:2770124

**KEYWORDS** GSS.

**SOURCE** human.

**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 487)  
**AUTHORS** Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden  
 ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.  
 and Venter,J.C.

**TITLE** Use of a random BAC End Sequence Database for Sequence-Ready Map  
 Building

**JOURNAL** Unpublished (1997)

**COMMENT** Other\_GSSs: CIT-HSP-2028F18.TR

Contact: Mark Adams

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC  
 end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)

Seq primer: M13-21

Class: BAC ends.

Location/Qualifiers

1..487

/organism="Homo sapiens"

## FEATURES

## source

## BASE COUNT

## ORIGIN

141 a 121 c 85 g 154 t

others

/db\_xref="GDB:7048076"  
 /db\_xref="taxon:9606"  
 /clone="2028F18"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /note="vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
 HindIII"

**BASE COUNT** 160 a 101 c 100 g 126 t

## ORIGIN

**Query Match** 80.0%; Score 16.8; DB 13; Length 487;  
**Best Local Similarity** 90.0%; Pred. No. 4.9e+02;  
**Matches** 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**Qy** 2 cttggactataatacttgac 21

**Db** 57 CTTGGACTGTAATACTTGTGTC 76

**RESULT** 10

## LOCUS

## DEFINITION

HS\_5374\_Al\_H01.T7A RPCI-11 Human Male BAC Library Homo sapiens  
 genomic clone Plate=950 Col=1 Row=O, DNA sequence.

**ACCESSION** A0666336

**VERSION** A0666336.1 GI:5174104

**KEYWORDS** GSS.

**SOURCE** human.

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 505)

**AUTHORS** Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.

**TITLE** Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome

**JOURNAL** Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

**MEDLINE** 99380589

**COMMENT** Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering.bac.htm>)  
 or from Resear h Genetics ([info@resgen.com](http://info@resgen.com)). BAC end Web Server:  
<http://www.htsc.washington.edu>  
 Plate: 950 row: O column: 1  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 505.  
 Location/Qualifiers  
 1..505  
 .organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=950 Col=1 Row=O"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /sex="male"

**FEATURES**

## source

1..505

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=950 Col=1 Row=O"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBACE3.6 vector at EcoRI sites"

**BASE COUNT** 141 a 121 c 85 g 154 t

**ORIGIN**

```

Query Match      80.0%; Score 16.8; DB 13; Length 505;
Best Local Similarity 90.0%; Pred. No. 4.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactctgac 21
||||| ||| ||||| |||
Db 500 CTTGGAGTATTATCTTCAC 481

RESULT 11
A0836967      537 bp      DNA      GSS      30-AUG-1999
DEFINITION   HS_4553_A1_D09_T7A CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=4553 Col=17 Row=G, DNA sequence.
ACCESSION    A0836967
VERSION      A0836967.1 GI:5806929
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 537)
Maharasa,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE        Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE      99380589
COMMENT      Contact: Maharasa GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 4553 row: G column: 17
Seq primer: T7
Class: BAC ends
High quality sequence stop: 537.
Location/Qualifiers
1. .537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=4553 Col=17 Row=G"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH108"

BASE COUNT    154 a 128 c 115 g 137 t 3 others
ORIGIN

Query Match      80.0%; Score 16.8; DB 13; Length 537;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactctgac 21
||||| ||| ||||| |||
Db 438 CTTGGACTAGTACTCTTAC 457

RESULT 12
BE846195      600 bp      mRNA      EST      25-SEP-2000
DEFINITION   BE846195 232148 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION    BE846195
VERSION      BE846195.1 GI:10283019
KEYWORDS     EST.
SOURCE       cow.

Query Match      80.0%; Score 16.8; DB 11; Length 600;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactctgac 21
||||| ||| ||||| |||
Db 532 CTTGGACTCTATCTCTGAC 551

RESULT 13
BH132866/c     854 bp      DNA      GSS      07-AUG-2001
LOCUS         ENTNI55TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION     BH132866
VERSION       BH132866.1 GI:15091927
KEYWORDS      GSS.
SOURCE        Entamoeba histolytica.
ORGANISM      Entamoeba histolytica.
REFERENCE     Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 854)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
JOURNAL       Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: enta@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library

```

Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 18  
 High quality sequence stop: 742.  
 Location/Qualifiers

## FEATURES

source

1. .854  
 /organism="Entamoeba histolytica"  
 /strain="HM1:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHO51; Site\_1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Borell, Oxford University Press, 1999)."

BASE COUNT 251 a 100 c 95 g 408 t

## ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 854;  
 Best Local Similarity 90.0%; Pred. No. 5.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcttgactataacttga 20  
 ||||| ||||| ||||| |||||  
 Db 143 GCTTGGACTAAATATTGA 124

## RESULT 14

AZ670679/c ENTHD34TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 LOCUS 968 bp DNA GSS 14-DEC-2000  
 DEFINITION genomic, DNA sequence.  
 ACCESSION AZ670679  
 VERSION AZ670679.1 GI:11807825  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.  
 Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 968)  
 Loftus, B., Van Aken, S. and Fraser, C.  
 Determination of clone end sequences from Entamoeba histolytica  
 HM1:IMSS sheared DNA library  
 TITLE HM1:IMSS sheared DNA library  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: enta@tigr.org  
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared  
 DNA library

Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 94  
 High quality sequence stop: 406.  
 Location/Qualifiers

## FEATURES

source

1. .968  
 /organism="Entamoeba histolytica"  
 /strain="HM1:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHO51; Site\_1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Borell, Oxford University Press, 1999)."

BASE COUNT 254 a 127 c 210 g 377 t

## ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 968;  
 Best Local Similarity 90.0%; Pred. No. 5.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcttgactataacttga 20  
 ||||| ||||| ||||| |||||  
 Db 363 GCTTGGACTAAATATTGA 344

## RESULT 15

BB103479 301 bp mRNA EST 27-JUN-2000  
 BB103479 RIKEN full-length enriched, 12 days embryo, embryonic body  
 between diaphragm region and neck Mus musculus cDNA clone  
 9430087K10.3, mRNA sequence.

ACCESSION BB103479  
 VERSION BB103479.1 GI:8756047  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 301)  
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci  
 P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata  
 Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,  
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya  
 T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,  
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino  
 M., Muramatsu, M. and Hayashizaki, Y.

## REFERENCE

## AUTHORS

RIKEN Mouse ESTs (Konno, H., et al.)  
 Unpublished (2000)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki  
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Thermostabilization and thermoactivation of thermolabile enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

## TITLE

## JOURNAL

## COMMENT

19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source  
Location/Qualifiers  
1. .301  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="9430087K10"  
/clone.lib="RIKEN full-length enriched, 12 days embryo,  
embryonic body between diaphragm region and neck"  
/tissue\_type="embryonic body between diaphragm region and  
neck"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN, Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 20.0 and subtraction to Rot = 370.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATCTCGAGTTAATTAATTAATCCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI."

BASE COUNT 75 a 54 c 48 g 124 t  
ORIGIN

Query Match 78.1%; Score 16.4; DB 10; Length 301;  
Best Local Similarity 94.4%; Pred. No. 7.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ttggactataacttga 20  
| | | | | | | | | |  
Db 96 TTGGACTTTAATACTTGA 113

Search completed: February 25, 2002, 17:21:20  
Job time: 16173 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:02:56 ; Search time 2331.3 Seconds  
(without alignments)  
155.680 Million cell updates/sec

Title: US-09-698-903B-13

Perfect score: 22

Sequence: 1 aacgagtgctcagctagaccagc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*\*

1: gb\_ba:\*\*

2: gb\_htg:\*\*

3: gb\_in:\*\*

4: gb\_om:\*\*

5: gb\_ov:\*\*

6: gb\_pat:\*\*

7: gb\_ph:\*\*

8: gb\_pl:\*\*

9: gb\_pr:\*\*

10: gb\_ro:\*\*

11: gb\_sts:\*\*

12: gb\_sy:\*\*

13: gb\_un:\*\*

14: gb\_vl:\*\*

15: em\_ba:\*\*

16: em\_fun:\*\*

17: em\_hum:\*\*

18: em\_in:\*\*

19: em\_om:\*\*

20: em\_or:\*\*

21: em\_ov:\*\*

22: em\_pat:\*\*

23: em\_ph:\*\*

24: em\_pl:\*\*

25: em\_ro:\*\*

26: em\_sts:\*\*

27: em\_sy:\*\*

28: em\_un:\*\*

29: em\_vl:\*\*

30: em\_htgo\_hum:\*\*

31: em\_htgo\_inv:\*\*

32: em\_htgo\_rod:\*\*

33: em\_htg\_hum:\*\*

34: em\_htg\_inv:\*\*

35: em\_htg\_rod:\*\*

36: em\_htg\_other:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	22	100.0	22	6	AX127760	Sequence
2	22	100.0	22	6	AX172481	Sequence
3	22	100.0	1624	8	BNACRUC	MI6860 Rapeseed cr
4	22	100.0	3113	6	I24540	Sequence 20
5	22	100.0	3113	6	I83673	Sequence 2
6	22	100.0	3113	8	BNCRUA	X14555 Brassica na
7	22	100.0	3198	8	BNC1G	X59294 B.napus BnC
8	19.4	88.2	10959	9	AL139402	Human DNA
9	18.8	85.5	2346	8	ATCRAL	AL1312 Arabidopsis
10	18.8	85.5	2346	8	ATCRALAA	M37247 A.thaliana
11	18.8	85.5	3081	8	BNC2G	X59295 B.napus BnC
12	18.8	85.5	84544	8	AB005239	Arabidops
13	18.8	85.5	167645	2	AC007864	Trypanoso
14	18	81.8	639	2	AC007860	AC007864
15	18	81.8	957	2	AC039235	AC007860
16	18	81.8	1083	2	AC039716	AC039235
17	17.8	80.9	3394	9	HSU66406	AC039716
18	17.8	80.9	160457	2	AC008049	U66406 Human putat
19	17.8	80.9	161416	2	AC087388	AC008049 Homo sapi
20	17.8	80.9	170871	2	AC084135	AC087388 Homo sapi
21	17.4	79.1	153865	8	AF002744	AC084135 Homo sapi
22	17.2	78.2	180	8	BNCRU4SUB	AP002744 Oryza sat
23	17.2	78.2	1211	8	ATHAT2S4	X57851 B.napus cru
24	17.2	78.2	2876	8	AT2SALBGA	M22033 A.thaliana
25	17.2	78.2	18177	2	AC020325	Z24744 A.thaliana
26	17.2	78.2	45980	8	AT24A18	AL035580 Arabidops
27	17.2	78.2	63855	2	AC016794	AL035580 Arabidops
28	17.2	78.2	77862	3	AC005465	AC016794 Drosophill
29	17.2	78.2	112664	9	AC020719	AC005465 Drosophill
30	17.2	78.2	176732	2	AC016870	AC020719 Homo sapi
31	17.2	78.2	178257	2	AC011781	AC016870 Homo sapi
32	17.2	78.2	198372	8	ATCHRIV66	AC011781 Homo sapi
33	17.2	78.2	249001	3	AE003824	AL161566 Arabidops
34	17	77.3	154732	9	AC011700	AE003824 Drosophill
35	17	77.3	184055	2	AL359648	AC011700 Homo sapi
36	16.8	76.4	1692	6	AX030716	AL359648 Homo sapi
37	16.8	76.4	79215	2	AC025522	AX030716 Sequence
38	16.8	76.4	172350	2	AC079234	AC025522 Homo sapi
39	16.8	76.4	174547	9	AC006445	AC079234 Homo sapi
40	16.8	76.4	189126	2	AC021132	AC006445 Homo sapi
41	16.8	76.4	190284	2	AC090572	AC021132 Homo sapi
42	16.8	76.4	210008	2	AC016974	AC090572 Homo sapi
43	16.8	76.4	210204	2	AC026358	AC016974 Homo sapi
44	16.4	74.5	110000	2	AC087799_2	AC026358 Homo sapi
45	16.4	74.5	161700	2	AL603793	Continuation (3 of AL603793 Mus muscu

ALIGNMENTS

RESULT 1						
AX127760	AX127760	Sequence 13 from Patent WO0131042.	22 bp	DNA	PAT	15-MAY-2001
LOCUS	AX127760	Sequence 13 from Patent WO0131042.				
DEFINITION	AX127760					
ACCESSION	AX127760.1	GI:14134407				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						

synthetic construct.  
artificial construct  
1 (bases 1 to 22)  
Weston,B. and de Beuckeleer,M.  
Male-sterile brassica plants and methods for producing same  
Patent: WO 0131042-A 13 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
Location/Qualifiers  
1..22  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer CV27"

```
BASE COUNT      7 a      6 c      6 g      3 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
|||||
Db 1 AACGAGTGTCTAGCTAGACCAGC 22

RESULT 2
AX172481
LOCUS      AX172481      22 bp      DNA      PAT      03-JUL-2001
DEFINITION Sequence 42 from Patent WO0141558.
ACCESSION AX172481
VERSION
KEYWORDS
SOURCE      synthetic construct.
ORGANISM      artificial sequence.
REFERENCE      1 (bases 1 to 22)
AUTHORS      de Both, G. and de Beuckeleer, M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL      Patent: WO 0141558-A 42 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES
source      1..22
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer BNA05"

BASE COUNT      7 a      6 c      6 g      3 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
|||||
Db 1 AACGAGTGTCTAGCTAGACCAGC 22

RESULT 3
BNACRUC
LOCUS      BNACRUC      1624 bp      mRNA      PLN      27-APR-1993
DEFINITION Rapeseed cruciferin mRNA, complete cds.
ACCESSION M16860
VERSION
KEYWORDS      12S storage protein; cruciferin.
SOURCE      Rapeseed (cultivar Tower) embryo (26-day post anthesis), cDNA to mRNA, clone pC1.
ORGANISM      Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE      1 (bases 1 to 1624)
AUTHORS      Simon, A.E., Tenbarger, K.M., Scofield, S.R., Finkelstein, R.R. and Crouch, M.L.
TITLE      Nucleotide sequence of a cDNA clone of Brassica napus 12S storage protein shows homology with legumin from Pisum sativum
JOURNAL      Plant Mol. Biol. 5, 191-201 (1985)
FEATURES
source      1..1624
/organism="Brassica napus"
/db_xref="taxon:3708"
<1..1624
/product="CRUC mRNA"
mRNA      17..85
sig_peptide
/note="cruciferin signal peptide"
```

## CDS

```
17..1483
/note="cruciferin precursor"
/codon_start=1
/protein_id="AAA32988.1"
/db_xref="GI:167134"
/translation="MARLSSLLSFLALIFLHGSTAQQFPNECQLDQLNALEPSSHVL
KAEAGRIEVDHHPOLRCGSFVRYIIIESKGLYLPSPFFSTAKLSFVAKGGLMGRV
VLCAETFDQSVFPSPGSGSPFGEGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOG
FRDMHOKVEHRTGTDTIATHPGVAQWPFYNDGNQPLVIYVSLASHONQIDNRPFY
LAGNPGQVWIEGRQQPKNILGFTPEVLAKAFKIDVRTAQQLNQODNRGNLIIR
VOGPFVIRPPLRSQRPOEVEVNGLETICSARCTDNLDPPSNADYKPKQGYISTLNS
YDLPIRLRLSALRGSIQNMVLPQNMNANAVLYTVDGEAHVQVVDNGDRVFDG
QVSGQLLSIQGFSVVKRATSEQFWIEFKTNNAAQINTLAGRTSVLRGLPLEVISN
GYQISLEEARVKENTITETLTHSSGPASYGGPRKADA"
mat_peptide      86..910
/note="cruciferin alpha subunit"
mat_peptide      911..1480
/note="cruciferin beta subunit"
BASE COUNT      457 a      425 c      387 g      355 t
ORIGIN      70 bp upstream of NcoI site.

Query Match      100.0%; Score 22; DB 8; Length 1624;
Best Local Similarity 100.0%; Pred. No. 0.098;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
|||||
Db 98 AACGAGTGTCTAGCTAGACCAGC 119

RESULT 4
I24540
LOCUS      I24540      3113 bp      DNA      PAT      07-OCT-1996
DEFINITION Sequence 20 from patent US 5543576.
ACCESSION I24540
VERSION I24540.1 GI:1604410
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 3113)
AUTHORS      van Ooijen, A.J., Rietveld, K., Hoekema, A., Pen, J., Sijmons, P.C.,
Verwoerd, T.C. and Quax, W.J.
TITLE      Production of enzymes in seeds and their use
JOURNAL      Patent: US 5543576-A 20 06-AUG-1996;
FEATURES
source      1..3113
/organism="unknown"
BASE COUNT      961 a      685 c      586 g      881 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 3113;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
|||||
Db 790 AACGAGTGTCTAGCTAGACCAGC 811

RESULT 5
I83673
LOCUS      I83673      3113 bp      DNA      PAT      10-AUG-1998
DEFINITION Sequence 2 from patent US 5714474.
ACCESSION I83673
VERSION I83673.1 GI:3407203
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 3113)
```

AUTHORS Van Ooljen,A.J.J., Rietveld,K., Hoekema,A., Pen,J.,  
Sijmons,P.Christian, Verwoerd,T.Cornelis and Quax,W.Johannes.  
TITLE Production of enzymes in seeds and their use  
JOURNAL Patent: US 5714474-A 2 03-FEB-1998;  
FEATURES Location/Qualifiers  
source  
1..3113  
/organism="unknown"  
BASE COUNT 961 a 684 c 586 g 882 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
|||||  
Db 790 AACGAGTGTCTAGACCAGC 811

RESULT 6  
BNCRUA BNCRUA 3113 bp DNA PLN 10-FEB-1999  
LOCUS Brassica napus cruA gene for cruciferin.  
DEFINITION X14555  
ACCESSION X14555  
VERSION X14555.1 GI:117810  
KEYWORDS cruA gene; cruciferin; seed storage protein.  
SOURCE rape.  
ORGANISM Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 3113)  
AUTHORS Ryan,A.J.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAR-1989) Ryan A.J., Department of Biological  
Sciences, University of Durham, South Road, Durham, DH1 3LE, United  
Kingdom  
2 (bases 1 to 3113)  
AUTHORS Ryan,A.J., Royal,C.L., Hutchinson,J. and Shaw,C.H.  
TITLE Genomic sequence of a 12S seed storage protein from oilseed rape  
(Brassica napus c.v. jet neuf)  
JOURNAL Nucleic Acids Res. 17 (9), 3584 (1989)  
MEDLINE 89263796  
FEATURES Location/Qualifiers  
source  
1..3113  
/organism="Brassica napus"  
/strain="Jet Neuf"  
/db\_xref="taxon:3708"  
/tissue\_type="leaf"  
644..650  
/note="pot. TATA-box"  
promoter 580..3010  
precursor\_rna 580..3010  
/note="primary transcript"  
CDS join(709..991,1220..1581,2049..2870)  
/codon\_start=1  
/product="cruciferin"  
/protein\_id="CAA32692.1"  
/db\_xref="GI:17811"  
/db\_xref="SWISS-PROT:P11090"  
/translation="MARLSLLSFLALLFLHGSTAQQFPNECQLDQLNALEPSSHVL  
KAEAGRIEVDHHPQLRCSGVSFVRYIESKGLYLPSFFSTAKLSFVAKGEGLMGRV  
VLCAEFTQDSSVFQSGSPGEGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOG  
FRDMHQKVEHRTGDTATHPGVAQWYNDGNQPLVIVSLDLASHQNLDRNRPFY  
LAGNPGQGWIEGREQQPNKILNGEFTPEVLAKAFKIDVTAQQLNQDNRGNILR  
VQGFPSVIRPLRSRQFEVNGLEETICSARCTNDLDDPSNADYKPOLQYISTLS  
YDLPILRLSALRSIRQANMLPQWNAVANAVLYVTGDEAHVQVWVNDGDRVFDG  
GYQSGQLLSIPOGFSVVKRATSEQFRFETKTNAAQINTLAGRTSVLRGLPLEVIS  
GYQISLEARRVKFTIETTLTHSSGPASVGGPRKADA"  
992..1219  
/note="intron I"  
intron 1582..2048  
/note="intron II"

AUTHORS Van Ooljen,A.J.J., Rietveld,K., Hoekema,A., Pen,J.,  
Sijmons,P.Christian, Verwoerd,T.Cornelis and Quax,W.Johannes.  
TITLE Production of enzymes in seeds and their use  
JOURNAL Patent: US 5714474-A 2 03-FEB-1998;  
FEATURES Location/Qualifiers  
source  
1..3113  
/organism="unknown"  
BASE COUNT 961 a 684 c 586 g 882 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
|||||  
Db 790 AACGAGTGTCTAGACCAGC 811

RESULT 7  
BNC1G BNC1G 3198 bp DNA PLN 04-APR-1995  
LOCUS B.napus Bnc1 gene for cruciferin storage protein.  
DEFINITION X59294  
ACCESSION X59294  
VERSION X59294.1 GI:17790  
KEYWORDS cruciferin; cruciferin storage protein.  
SOURCE rape.  
ORGANISM Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 3198)  
AUTHORS Breen,J.P. and Crouch,M.L.  
TITLE Molecular analysis of a cruciferin storage protein gene family of  
Brassica napus  
JOURNAL Plant Mol. Biol. 19 (6), 1049-1055 (1992)  
MEDLINE 92379259  
REFERENCE 2 (bases 1 to 3198)  
AUTHORS Breen,J.P.  
TITLE Direct Submission  
JOURNAL Submitted (24-APR-1991) Breen J.P., Jordan Hall, Indiana  
University, Bloomington, IN 47401, USA  
COMMENT See also M16860 & X59295 (for Bnc2 gene).  
FEATURES Location/Qualifiers  
source  
1..3198  
/organism="Brassica napus"  
/strain="C.v. Tower"  
/db\_xref="taxon:3708"  
/cell\_line="ED8767"  
/clone\_lib="Ch4a"  
/clone="lambda Bnc1"  
join(<709..991,1220..1584,2069..2488,2653..>3057)  
/gene="Bnc1"  
709..3057  
/gene="Bnc1"  
join(709..991,1220..1584,2069..2488,2653..3057)  
/gene="Bnc1"  
/codon\_start=1  
/product="cruciferin storage protein"  
/protein\_id="CAA41984.1"  
/db\_xref="GI:762919"  
/db\_xref="SWISS-PROT:P33523"  
/translation="MARLSLLSFLALLFLHGSTAQQFPNECQLDQLNALEPSSHVL  
KAEAGRIEVDHHPQLRCSGVSFVRYIESKGLYLPSFFSTAKLSFVAKGEGLMGRV  
VLCAEFTQDSSVFQSGSPGEGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOG  
FRDMHQKVEHRTGDTATHPGVAQWYNDGNQPLVIVSLDLASHQNLDRNRPFY  
LAGNPGQGWIEGREQQPNKILNGEFTPEVLAKAFKIDVTAQQLNQDNRGNILR  
VQGFPSVIRPLRSRQFEVNGLEETICSARCTNDLDDPSNADYKPOLQYISTLS  
YDLPILRLSALRSIRQANMLPQWNAVANAVLYVTGDEAHVQVWVNDGDRVFDG  
GYQSGQLLSIPOGFSVVKRATSEQFRFETKTNAAQINTLAGRTSVLRGLPLEVI  
NGYQISLEARRVKFTIETTLTHSSGPASVGGPRKADA"  
709..991  
/gene="Bnc1"  
/number=1  
992..1219

```

/gene="BnCl"
/number=1
1220..1584
/gene="BnCl"
/number=2
1585..2068
/gene="BnCl"
/number=2
2069..2488
/gene="BnCl"
/number=3
2489..2652
/gene="BnCl"
/number=3
2654..>3057
/gene="BnCl"
/number=4
3102..3107
polyA_signal
polyA_signal
3176..3181
BASE COUNT 992 a 690 c 600 g 916 t
ORIGIN

```

```

Query Match 100.0%; Score 22; DB 8; Length 3198;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22
|||||
Db 790 AACGAGTGTCTAGCTAGACCAGC 811

```

```

RESULT 8
AL139402
LOCUS
DEFINITION Human DNA sequence from clone RP4-808F24 on chromosome
Xp11.23-11.4. Contains STSS and GSSs, complete sequence.
ACCESSION AL139402.11 GI:9944136
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 10959)
Clark, S.
Direct Submission
Submitted (24-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 29, 2000 this sequence version replaced gi:9926454.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; SW., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep/
This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP4-808F24 is from the library RPCI-4 constructed at the Roswell

```

Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>  
 VECTOR: pCYPAC2  
 IMPORTANT: This sequence is not the entire insert of clone RP4-808F24 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP4-774G10 is at 10860 in this sequence.  
 The true right end of clone RP6-227L5 is at 100 in this sequence.

## FEATURES

```

source
Location/Qualifiers
1..10959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11.23-11.4"
/clone="RP4-808F24"
/clone_lib="RPCI-4"
/complement(1..87)
/note="match: STS: Em:HS227L5T"
2314..2355
/note="21 copies 2 mer gg 76% conserved"
2339..2342
/note="unidirectional dGTP"
complement(2797..3221)
/note="match: GSS: Em:B83505"
5384..5631
/note="L1MB8 repeat: matches 5924..6175 of consensus"
5682..5791
/note="L1MB8 repeat: matches 5654..5773 of consensus"
6792..6940
/note="LTR37A repeat: matches 268..420 of consensus"
6941..6978
/note="19 copies 2 mer ta 92% conserved"
6980..7218
/note="LTR37A repeat: matches 47..282 of consensus"
7440..7486
/note="LTR37A repeat: matches 1..47 of consensus"
8540..8922
/note="MSTA repeat: matches 1..426 of consensus"
complement(8828..9278)
/note="match: GSS: Em:AQ223257"
complement(8948..9266)
/note="match: GSS: Em:B53441"
9300..9420
/note="MER5B repeat: matches 1..128 of consensus"
complement(9616..10040)
/note="match: GSS: Em:AQ369714"
10048..10487
/note="L1MB3 repeat: matches 5455..5901 of consensus"
10509..10820
/note="LTR26 repeat: matches 291..600 of consensus"
10847..10958
/note="LTR26 repeat: matches 201..320 of consensus"
BASE COUNT 3380 a 2058 c 2220 g 3301 t
ORIGIN

```

```

Query Match 88.2%; Score 19.4; DB 9; Length 10959;
Best Local Similarity 95.2%; Pred. No. 3.7;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 aacgagtgtcagctagaccag 21
|||||
Db 7197 AAGGAGTGTCTAGCTAGACCAG 7217

```

## RESULT 9

```

ATCCAL
LOCUS ATCCAL 2346 bp DNA PLN 18-MAY-1995
DEFINITION Arabidopsis CRAL gene for L2S seed storage protein.
ACCESSION X14312
VERSION X14312.1 GI:16231
KEYWORDS CRAL gene; seed storage protein; storage protein.
SOURCE thale cress.

```



ORGANISM	Arabidopsis thaliana	ORGANISM	eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 2346)	REFERENCE	1 (bases 1 to 2346)
AUTHORS	Pang,P.P., Pruitt,R.E. and Meyerowitz,E.M.	AUTHORS	Pang,P.P., Pruitt,R.E. and Meyerowitz,E.M.
TITLE	Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana	TITLE	Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana
JOURNAL	Plant Mol. Biol. 11, 805-820 (1988)	JOURNAL	Plant Mol. Biol. 11, 805-820 (1988)
FEATURES	Location/Qualifiers	FEATURES	Location/Qualifiers
source	1..2346	source	1..2346
	/organism="Arabidopsis thaliana"		/organism="Arabidopsis thaliana"
	/cultivar="Landsberg erecta"		/db_xref="taxon:3702"
	/db_xref="taxon:3702"		/protein_id="AAA32777.1"
promoter	138..147		/db_xref="GI:166676"
exon	196..496		/translation="MARVSSLSFCLTLLILFHGYAAQGGQGGQFPNECOLDLNAL EPSHVLKSEAGRIEVDHAPQLRCSGVSEAFYILIESKGLYLPSFFNTAKLSFVAKGR GLMGKVI PGCAETFDQSEFPREGGQSQRFDHOKVHEIRSGDTIATPGVAQM FYNDGQPLIVSVFDLASHQNDLRNRPFYLAGNPPQGVWLOGREQPQKNIFNG FGPEVIAQLKIDIQTAQLQNDNRNIVRVQGFVIRPPLRGORPQEEEEEGR HGRHNGLEETICSACTDNLDDPSRADYKPOLGYISTLNSYDLPLIRLISALRG SIRONAMVLQWNAANAILIETDGEAQIOIVNDGNRVFDGQVSGQLTAVPGFSV VKRATSNRFQWVEKTNANAOINTLAGRTSVLRGLPLEVITNGFQISPEEARVKFNT LETTLTHSSGPASYGRPRVAAA"
CDS	join(196..496,623..888,985..1431,1547..1951)	exon	<196..496
	/product="12S seed storage protein"		/note="12S storage protein CRA1"
	/protein_id="CAA32493.1"	intron	/number=1
	/db_xref="GI:808936"		497..622
	/translation="MARVSSLSFCLTLLILFHGYAAQGGQGGQFPNECOLDLNAL EPSHVLKSEAGRIEVDHAPQLRCSGVSEAFYILIESKGLYLPSFFNTAKLSFVAKGR GLMGKVI PGCAETFDQSEFPREGGQSQRFDHOKVHEIRSGDTIATPGVAQM FYNDGQPLIVSVFDLASHQNDLRNRPFYLAGNPPQGVWLOGREQPQKNIFNG FGPEVIAQLKIDIQTAQLQNDNRNIVRVQGFVIRPPLRGORPQEEEEEGR HGRHNGLEETICSACTDNLDDPSRADYKPOLGYISTLNSYDLPLIRLISALRG SIRONAMVLQWNAANAILIETDGEAQIOIVNDGNRVFDGQVSGQLTAVPGFSV VKRATSNRFQWVEKTNANAOINTLAGRTSVLRGLPLEVITNGFQISPEEARVKFNT LETTLTHSSGPASYGRPRVAAA"	exon	/number=2
	497..622		623..888
exon	623..888	exon	/number=2
intron	889..984	intron	889..984
exon	985..1431	exon	985..1431
intron	1432..1546	intron	1432..1546
exon	1547..1951	exon	1547..1951
polyA_signal	2015..2020		/note="12S storage protein CRA1 intron B"
polyA_signal	2085..2090		/note="12S storage protein CRA1 intron C"
BASE COUNT	714 a 528 c 472 g 632 t	BASE COUNT	714 a 528 c 472 g 632 t
ORIGIN		ORIGIN	
Query Match	85.5%; Score 18.8; DB 8; Length 2346;	Query Match	85.5%; Score 18.8; DB 8; Length 2346;
Best Local Similarity	90.9%; Pred. No. 7.8;	Best Local Similarity	90.9%; Pred. No. 7.8;
Matches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	Matches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 aacgagtgtcagctagaccagc 22 	Qy	1 aacgagtgtcagctagaccagc 22 
Db	295 AACGAGTGCCAGCTCGACCAGC 316	Db	295 AACGAGTGCCAGCTCGACCAGC 316
RESULT	11	RESULT	11
BNC2G		BNC2G	
LOCUS	BNC2G 3081 bp DNA	LOCUS	BNC2G 3081 bp DNA
DEFINITION	B.napus Bnc2 gene for cruciferin storage protein.	DEFINITION	B.napus Bnc2 gene for cruciferin storage protein.
ACCESSION	X59295	ACCESSION	X59295
VERSION	X59295.1 GI:17791	VERSION	X59295.1 GI:17791
KEYWORDS	cruciferin; cruciferin storage protein.	KEYWORDS	cruciferin; cruciferin storage protein.
SOURCE	rape.	SOURCE	rape.
ORGANISM	Brassica napus	ORGANISM	Brassica napus
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 3081)		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid II; Brassicales; Brassicaceae; Brassica. 1 (bases 1 to 3081)
REFERENCE	1 (bases 1 to 3081)	REFERENCE	1 (bases 1 to 3081)
AUTHORS	Breen,J.P. and Crouch,M.L.	AUTHORS	Breen,J.P. and Crouch,M.L.
TITLE	Molecular analysis of a cruciferin storage protein gene family of Brassica napus	TITLE	Molecular analysis of a cruciferin storage protein gene family of Brassica napus
JOURNAL	Plant Mol. Biol. 19 (6), 1049-1055 (1992)	JOURNAL	Plant Mol. Biol. 19 (6), 1049-1055 (1992)
MEDLINE	92379259	MEDLINE	92379259
REFERENCE	2 (bases 1 to 3081)	REFERENCE	2 (bases 1 to 3081)
AUTHORS	Breen,J.P.	AUTHORS	Breen,J.P.
TITLE	Direct Submission	TITLE	Direct Submission



```
CDS
10646..10686,10782..10857))
/note="gene_id:MLN1.2"
/codon_start=1
/evidence="not_experimental
/product="casein kinase I"
/protein_id="BAB10977.1"
/db_xref="GI:9759511"
/translation="MDLVIGGKFKLGGKIGSGSGELYLVNVTGGEVAVKLENVKT
KHPQLHYESKLYMLLQGGSGIPNKFVGGEDSVVMYIDLIGPSLEDLFNCNKLKTL
KTVLMLADQLNVRFEFMHTGRFLHRDKLPDNFLMGLGRKANQVYIIDFGLGKKYRDQLQ
TKHPIYPRKNKLTGTARYASVNTHLGVEQSRDDLESGLGYVLMYFLKGLSPWQGLKA
GTTKKQYDRLSKEKVSPIEVLCNQPSEFVSYPHYCRSLREFDDKPDYSYLKRLFRDL
FIREGYQFYFDWTVLYKYPQIGSSGSSSTRHHTTAKGFGNADPIERGERILCKET
SRKVAFAFARRHTTSSPPDRSRNSDDGPFSSKQTHGDSERAKNSSSRRASS
TRKVAASSRRSPAGPTTSSRLVSSGGGSGSGNGRPSTSRVQAQYSESKTL
SFSRATAARNTREDQLRSFELLSLRK"
complement(join(12794..13054,13133..13219,13304..13440,
13520..13571,13656..13816,14094..14244))
/note="gb|AAD20643.1
gene_id:MLN1.3"
/codon_start=1
/evidence="not_experimental
/product="NBD-like protein"
/protein_id="BAB10978.1"
/db_xref="GI:9759512"
/translation="MAEKNASAVDGAIRVSGMQFSYDVQDPFFDFNLDLPAGSRCLL
VGANGSTKTLKLAGHMGVGNVVOVLDRSAFHDTLVCSDGLSYLGGSWSKTAG
SAGDIPQGGSAFHMIFFGVGIDPPREKLLDLDIDNLOWRMHKVSDGQRRRVQICM
ALLHPFKVLLDEVTVLDVVARMDLLEFPKECEQRGATYVYATIFPDGLTWSHL
AYINGVKLUSALDEIKDLKTLSPNLVSVEAWLRSETKVKTKKPPVVTSPMWSR
OMAYR"
complement(join(15032..15436,15552..15998,16129..16394,
16518..16818))
/note="gene_id:MLN1.4"
/codon_start=1
/evidence="not_experimental
/product="legumin-like protein"
/protein_id="BAB10979.1"
/db_xref="GI:9759513"
/translation="MARVSSLLSFLCLLLILFHGYAAQGGQGGQFNECQLDOLNAL
EPHVLKSEAGRIEVDHHPQLRCGVSFARYILIESKGLYLPSEFFNTAKLSFVAKGR
GLMKVTPGCAETFDQDSEFPQREGQSFQRMHQKHVEHRSIGDTIATTPGVAQM
FYNDGQPLVYVFDIAHQNDLRNRPFYLAGNQPOGVQWLRQEQPKNFING
FGVEYAIQKIDIQALQONQDNRGNIVRVQPGFVRPRLRGQRPQEEEEEER
HGRGNLEETICSACTDNLDDPSRADYKPOLGYITSLNSYDPLTLRFLRLSALRG
SIQNNAWLPQMANANAILYVTDGEAQIOIVDNGNRVEDGOVSOGOLAVPQGFV
VKRATSRFQWFEKTNANAQINTLAGRTSVLRGLPLEVINGQISPEARRVKFNT
LETLTHSSGPASTGRPRVAAA"
19700..20443
/note="contains similarity to surface protein
gene_id:MLN1.5"
/codon_start=1
/evidence="not_experimental
/protein_id="BAB10980.1"
/db_xref="GI:9759514"
/translation="MATPTLLLLLTAVFLSTETAQRAAPAGPAGPINITALEKG
GQFVTLRLNTQTQIGNQINQINSSEGTVLAPTNAFQNLKPGTLNKLSPDDQVK
LILLYHSPKPYTLEDLLSVNPNVTOASGRDVGGYGLNTGTGGQNVNVTGGVYETRL
STSLQRGPLAVYVDVLLPEEMFGRKISPMAPPKSPDVSDDSSESKKAAAPS
ESKSGSGEMTGLGLGLVLCLELL"
join(21063..21237,21296..21408,21497..21667,21741..21953,
22037..22201)
/note="gene_id:MLN1.6"
/codon_start=1
/evidence="not_experimental
/product="prohibitin"
/protein_id="BAB10981.1"
/db_xref="GI:9759515"
/translation="MNVKVPNVPGSPALSALLKLVIGLGLYIGSSMNVVDGGRH
AIVNFTGKIDRVYPTGQNFKEKPLFERALYIDVRSPYVENSQTSGNDQLQVTVIGR
VLTRPMGRDLPETIRLTGQNGERVLSPILNETLKAVAQYNASHLITQREAVYRIR
KIYTERAKFNALDDVSTNLKFGKEFTAEIKKQVAQEAERAKFIVEAKQDDKS
AIIRAQGEAKSAQLIGQAIANNEAFITLRKIEAREIAQTIAKSANKVYLNSSDLLIS
```

```
CDS
KQ"
complement(join(22369..22845,22924..23036,23129..23230,
23307..23673))
/note="gene_id:MLN1.7
unknown protein"
/codon_start=1
/evidence="not_experimental
/protein_id="BAB10982.1"
/db_xref="GI:9759516"
/translation="MDSKSLAKSKRAHTLHHSSKSHSVHKPKVPGVSEKNPEKLQGNQ
TKSPVQSRVRSALPNWDYDELDAAEDSSILSHSDVIVPKSGADYLHLISEAQAE
KSVLENNLSSLDLLHDEFRRVGSMTSARCEGILSHMEDDNFVVEEDGSKVPG
FLSLNVLAKTLENVDLHERLYIDPLLPPELNTSQTKVRNEEPESHQAQNDPI
VPGESSVREAESLDQVKDILIDDESEKSAEADLDDLNSSEAHQPNPVASAS
GKSAFTLEDSLKSHSSTQFNKPGNPQDKIHMTGFNDVLDLLESTPVSITPQS
NOTSSKVLDDFDSWLDTI"
join(31663..31801,32230..32626,33221..34085)
/note="contains similarity to zinc finger protein ID1
gene_id:MLN1.8"
/codon_start=1
/evidence="not_experimental
/protein_id="BAB10983.1"
/db_xref="GI:9759517"
/translation="MTSEVLQTISSGSGFAQPSSTLDHDESINPLVKKKNLPG
NPDEAEVIALSPITLMATNRFCEVCQGFQDQNLQHRHNLPMKLKORTSKY
KRTVYVPEKTCVHHSHSALGDLTGKHKFCRKGKKTCCKAKRYAQSDMKAH
SKTCREYRCDCGTIFSRSDSFTHRAFCALAEETAKINAVSHLNGLAAGAPGV
NLNYOLMGTFIPLPQFPVPOPQPNHGHQFOPPTSSLSLAWGODIAPQOPQDY
DWFVGNAAACACIDNNNTHDEQITONNASLTMTTTLSPSLFSSDQPNANANSV
NMSATALLKAAEIGATSTTAANDPSTFLOSPLKSTDTQTTSYDSEKGFALFGSN
NNIGMSRSHDQIEENARNDTVASALDELQNTYPMKRRRYDGGEGVGGGQTRDFLG
VGVTLCHPSSINGWI"
38966..39670
/note="gene_id:MLN1.9
pir||T38261
similar to unknown protein"
/codon_start=1
/evidence="not_experimental
/db_xref="GI:9759518"
/translation="MKFTDPSVIDLTVNGTKLSIQDNGSMHVGTSVWPCSLILSKFA
ERWTLSSSTTSPNPAELFDFRRRGIELGTCGVAGMAFYLLGLTEIVLTDIAP
VMPALKNLKRNKTAGLSKTSIVYNNRNQISALPKPFDLVTAAVYVYTESVGOL
VTAMELVADDDAGVLLGYQIRSEADKLFWELCDIVFKIEKVPHEHLHSYAYEETDV
YIFRKVKKNEAESV"
join(41744..41855,41947..42360,42441..42520,42621..43001,
43096..43185,43267..43431,43518..43740,43847..43947,
44084..44170,44251..44403,44498..44653,44728..44937,
45026..45676,45764..46294,46387..46947,47129..47233,
47333..47444)
Query Match 85.5%; Score 18.8; DB 8; Length 84544;
Best Local Similarity 90.9%; Pred. No. 9.2;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 aacgagtgcagctagaccagc 22
||||||| ||||| |||||
Db 16719 AACGAGTCCGACTGCACCAGC 16698

RESULT 13
AC007864
LOCUS AC007864 167645 bp DNA HTG 17-JUL-2001
DEFINITION Trypanosoma brucei chromosome II clone RPC193-28H13, *** SEQUENCING
IN PROGRESS ***, 1 ordered pieces.
ACCESSION AC007864
VERSION AC007864.4 GI:14787195
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 167645)
AUTHORS El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K.,
```



Db 918 GAGTGTCTCAGCTAGACCAG 935

Search completed: February 25, 2002, 18:03:03  
Job time: 18591 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:41 ; Search time 716.55 seconds  
(without alignments)  
26.322 Million cell updates/sec

Title: US-09-698-903B-13

Perfect score: 22

Sequence: 1 aacgagtgcagtagaccagc 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	22	100.0	22	22	AAH25457	PCR primer for end
2	22	100.0	22	22	AAH25457	PCR primer B03, to
3	22	100.0	3113	12	AAQ13870	Cruciferin A gene.
C 4	17.2	78.2	20633	20	AAQ13213	Enterococcus faeca
C 5	16.8	76.4	1692	18	AAQ60974	NLERK2 cDNA 3' reg
6	16.2	73.6	353	22	AAQ15388	Pseudomonas aerugi
7	16.2	73.6	353	22	AAQ15493	Pseudomonas aerugi
8	15.8	71.8	716	22	AAH05632	Human cDNA clone (
9	15.8	71.8	850	22	AAH07552	Human cDNA sequenc
10	15.8	71.8	1790	22	AAH14119	Human cDNA sequenc
11	15.8	71.8	1929	22	AAH16554	Human cDNA sequenc

12	15.8	71.8	2096	22	AAH05091	Human secreted pro
C 13	15.8	71.8	10708	19	AAV69286	Sequence of mouse
14	15.6	70.9	163	22	AAI52779	Probe #21465 used
15	15.6	70.9	578	22	AAI39698	Probe #8384 used t
C 16	15.6	70.9	695	21	AAI2137	Aspergillus oryzae
C 17	15.6	70.9	1193	21	AAQ41375	Arabidopsis thalia
C 18	15.6	70.9	1236	21	AAQ40221	Arabidopsis thalia
C 19	15.6	70.9	1238	21	AAQ49331	Arabidopsis thalia
C 20	15.6	70.9	1239	21	AAQ49330	Arabidopsis thalia
C 21	15.6	70.9	1240	21	AAQ33542	Arabidopsis thalia
C 22	15.6	70.9	1968	21	AAQ6946	Human secreted pro
23	15.6	70.9	2105	22	AAI58985	Human polynucleoti
C 24	15.6	70.9	2224	22	AAI60771	Human polynucleoti
C 25	15.6	70.9	3192	21	AAQ49328	Arabidopsis thalia
C 26	15.6	70.9	8390	21	AAQ22283	BAC containing rep
C 27	15.6	70.9	90336	21	AAQ22289	BAC containing rep
C 28	15.4	70.0	307	21	AAQ43854	Human secreted exp
C 29	15.4	70.0	1256	21	AAZ60654	DNA encoding a kap
C 30	15.4	70.0	1330	16	AAQ75928	Mouse oploid recep
C 31	15.4	70.0	2518	21	AAZ60655	DNA encoding a kap
C 32	15.4	70.0	2600	16	AAQ90096	Mouse kappa-3 opio
C 33	15.4	70.0	2634	21	AAZ60653	DNA encoding a kap
C 34	15.4	70.0	4315	20	AAQ12938	Enterococcus faeca
C 35	15.2	69.1	191	22	AAI52777	Probe #23963 used
C 36	15.2	69.1	350	13	AAQ31549	MAD-3. Homo sapie
37	15.2	69.1	520	22	AAI42192	Probe #10878 used
38	15.2	69.1	657	15	AAQ54681	Rat proteosome RIN
39	15.2	69.1	754	22	AAH29516	Drosophila melanog
C 40	15.2	69.1	819	18	AAQ91841	Human inhibitory k
C 41	15.2	69.1	930	20	AAQ22769	Human SOCS14 cDNA
C 42	15.2	69.1	1148	22	AAQ26304	Pseudomonas sp Typ
C 43	15.2	69.1	1548	19	AAV55840	Human IkappaB poly
C 44	15.2	69.1	1550	13	AAQ31550	Encodes IkB NF-kap
C 45	15.2	69.1	1550	19	AAV41719	Human I-kappa-B-al

ALIGNMENTS

RESULT 1

AAH25457

ID AAH25457 standard; DNA; 22 BP.

XX AC AAH25457;

XX DT 22-AUG-2001 (first entry)

XX DE PCR primer for endogenous sequences in transgenic plants.

XX KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;

XX KW fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.

XX OS Synthetic.

XX PN WO200141558-A1.

XX PD 14-JUN-2001.

XX PF 06-DEC-2000; 2000WO-EP12872.

XX PR 08-DEC-1999; 99US-0457037.

XX PA (AVET ) AVENTIS CROPS SCIENCE NV.

XX PI De Both G, De Beuckeleer M;

XX DR WPI; 2001-381419/40.

XX PT Transgenic winter oilseed rape plants suited for producing hybrid seed

XX PT with improved qualities, comprises a male-sterility gene and fertility

XX PT restorer gene, integrated into the genome -

XX PS Example 5; Page 53; 98pp; English.

XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AA25457-58 were used to amplify endogenous sequences  
 CC from transgenic plants of the invention.  
 XX  
 SQ Sequence 22 BP; 7 A; 6 C; 6 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgctcagctagaccagc 22  
 |||||  
 Db 1 aacgagtgctcagctagaccagc 22

RESULT 2  
 AAD07002  
 ID AAD07002 standard; DNA; 22 BP.

AC AAD07002;

XX 06-AUG-2001 (first entry)

XX PCR primer B03, to recognise foreign DNA and flanking sequence of MS-B2.  
 DE MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; ss.  
 KW Brassica napus.

XX WO200131042-A2.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-EP10680.

XX 29-OCT-1999; 99US-0430497.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX Weston B, De Beuckeleer M;

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX Example 5; Page 33; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.

XX Sequence 22 BP; 7 A; 6 C; 6 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgctcagctagaccagc 22  
 |||||  
 Db 1 aacgagtgctcagctagaccagc 22

RESULT 3  
 AAQ13870  
 ID AAQ13870 standard; DNA; 3113 BP.

XX AAQ13870;

XX 09-DEC-1991 (first entry)

XX Cruciferin A gene.

XX Seed storage protein; crua; ss.

XX Brassica napus.

XX EP449376-A.

XX 02-OCT-1991.

XX 25-MAR-1991; 91EP-0200688.

XX 25-MAR-1991; 91EP-0200688.

XX 23-MAR-1990; 90US-0498561.

XX (KONN ) GIST-BROCADES NV.

XX Pen J, Stijmons PC, Van Ooyen AJJ, Rietveld K, Verwoerd TC;

XX Quax WJ;

XX WPI; 1991-289815/40.

XX Seeds contg. enhanced enzyme levels from transgenic plants - used  
 PT for catalysing reactions, increasing nutritional values or  
 PT treating digestive disorders.

XX Example; Fig 3; 38pp; English.

XX The DNA is the genomic sequence of the seed storage protein gene  
 CC cruciferin A (crua). It can be used in the prodn. of transgenic  
 CC plants expressing cruciferin in its seeds for use in an industrial  
 CC process. The seeds contg. the cruciferin can be used without the  
 CC need for first extracting and/or isolating the enzymes. The use of  
 CC seeds for the storage of cruciferin provides a stable vehicle which  
 CC is easily packaged and transported and easily handled during use.  
 CC See also AAQ13871-Q13877.

XX Sequence 3113 BP; 961 A; 685 C; 586 G; 881 T; 0 other;

Query Match 100.0%; Score 22; DB 12; Length 3113;  
 Best Local Similarity 100.0%; Pred. No. 0.09;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgctcagctagaccagc 22  
 |||||  
 Db 790 aacgagtgctcagctagaccagc 811

RESULT 4  
 AAX13213/c  
 ID AAX13213 standard; DNA; 20633 BP.

XX AAX13213;

XX 19-MAR-1999 (first entry)



```
XX DE Enterococcus faecalis genome contig SEQ ID NO:276.
XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;
XX KW vaccine; attenuation; computer readable medium; ds.
XX OS Enterococcus faecalis.
XX XX
XX PN WO9850555-A2.
XX PD 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US08985.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
XX PI Barash SC, Dillon PJ, Kunsch CA;
XX XX
XX DR WPI; 1999-045171/04.
XX XX
XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides
XX PT - used to develop products for the detection of Enterococcus and for
XX PT use in vaccines for prevention or attenuation of Enterococcus
XX PT infection.
XX XX
XX PS Claim 1; Page 1284-1294; 2084pp; English.
XX CC
XX CC A computer readable medium has been developed which has recorded on it
XX CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX CC AAX12938 to AAX13919 represent these nucleotide sequences which are
XX CC primary nucleotide sequences, also known as contigs. The computer-based
XX CC system can identify fragments of the Enterococcus faecalis genome with
XX CC commercial importance. The products can be used to detect the presence
XX CC of Enterococcus faecalis in samples. They can also be used for
XX CC diagnosing Enterococcal infection in an animal and monitoring
XX CC progression of disease, and for identifying agents which can be used to
XX CC modulate the growth or pathogenicity of Enterococcus faecalis, or
XX CC another related organism, in vivo or in vitro. In particular the
XX CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX CC can be used in vaccines to prevent or attenuate an Enterococcal
XX CC infection.
XX XX
XX SQ Sequence 20633 BP; 5857 A; 4284 C; 3025 G; 7427 T; 40 other;

Query Match 78.2%; Score 17.2; DB 20; Length 20633;
Best Local Similarity 86.4%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
||||| ||| ||| |||||
Db 18911 AACGAGTATCATCTAAACCAGC 18890

RESULT 5
AAT60974/c
ID AAT60974 standard; cDNA; 1692 BP.
XX AC
XX AC AAT60974;
XX XX
XX DT 23-JUN-1997 (first entry)
XX XX
XX DE NLERK2 cDNA 3' region.
XX KW LERK; ligand for eph-related kinase; ERK; NLERK2;
XX KW receptor protein tyrosine kinase; cell proliferation;
XX KW cell differentiation; cell survival; nerve cell; ss.
XX XX
XX OS Homo sapiens.
```

```
XX PN WO9704091-A1.
XX XX
XX PD 06-FEB-1997.
XX XX
XX PF 19-JUL-1996; 96WO-AU00460.
XX XX
XX PR 05-FEB-1996; 96AU-0007890.
XX PR 20-JUL-1995; 95AU-0004263.
XX PR 27-NOV-1995; 95AU-0006847.
XX PR 22-DEC-1995; 95AU-0007299.
XX XX
XX PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX PI Nicola NA;
XX XX
XX DR WPI; 1997-132632/12.
XX XX
XX PT Nucleic acid mol. encoding ligand for eph-related kinase - useful
XX PT for treatment of, pref. neuronal, cells to increase survival,
XX PT proliferation and differentiation
XX XX
XX PS Claim 5; Page 43-45; 71pp; English.
XX CC
XX CC The 3' region (AAT60974) and 5' region (AAT60975) sequences were
XX CC detd. for a cDNA sequence (see also AAT60966) coding for NLERK2
XX CC (AAW10637), a novel human ligand for eph-related kinase (LERK). The
XX CC NLERK2 cDNA was obt'd. from a human foetal brain cDNA library using
XX CC probes (see also AAT60967-69) based on an expressed sequence tag
XX CC previously isolated on the basis of homology to conserved regions
XX CC of known LERKs (see also AAW10633-36).
XX XX
XX SQ Sequence 1692 BP; 316 A; 504 C; 385 G; 472 T; 15 other;

Query Match 76.4%; Score 16.8; DB 18; Length 1692;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acgagtgtcagctagaccagc 21
||||| ||||| ||||| |||||
Db 1434 ACCACTGTCTAGCTAGACCAG 1415

RESULT 6
AAF15388
ID AAF15388 standard; DNA; 353 BP.
XX AC
XX AC AAF15388;
XX XX
XX DT 09-MAR-2001 (first entry)
XX XX
XX DE Pseudomonas aeruginosa ssrA gene, SEQ ID NO:101.
XX XX
XX KW ssrA gene; tmRNA; bacterium; chloroplast; diatom; detection;
XX KW detection; identification; quantification; characterisation;
XX KW nucleic acid array; DNA chip; drug design; treatment monitoring;
XX KW contamination; ds.
XX XX
XX OS Pseudomonas aeruginosa.
XX XX
XX PN WO200070086-A1.
XX XX
XX PD 23-NOV-2000.
XX XX
XX PF 15-MAY-2000; 2000WO-IE000066.
XX PR 14-MAY-1999; 99WO-IE000043.
XX XX
XX PA (IRBI-) ENTERPRISE IRELAND T/A BIOTRESEARCH IRELA.
XX PA (UYN-) UNIV NAT IRELAND GALWAY.
XX XX
XX PI Barry TG, Smith TJ;
```

XX WPI; 2001-025025/03.  
 XX Use of ssrA gene, tmRNA, or fragments of them, as target regions in  
 PT probe assays for detection of prokaryotic or eukaryotic organisms, and  
 PT for determination of species -  
 XX  
 XX Disclosure; Page 55; 221pp; English.  
 XX  
 CC The invention relates to the ssrA gene or tmRNA, an RNA transcript  
 CC of the ssrA gene, or fragments thereof as target regions in a nucleic  
 CC acid probe assay for the detection and identification of prokaryotic  
 CC and/or eukaryotic organisms. The invention also relates to 38 novel ssrA  
 CC sequences and their tmRNA transcripts (claimed), and to pan-bacterial,  
 CC genus- and species-specific ssrA gene/tmRNA-directed probes and PCR  
 CC primers (claimed). tmRNA is a stable, high copy number RNA which is  
 CC found in all bacteria and is also found in chloroplasts and diatoms. It  
 CC has a dual function both as a tRNA and as an mRNA and is involved in  
 CC rescuing truncated mRNAs which have lost stop codons. ssrA genes and  
 CC tmRNA sequences can be used as target regions in nucleic acid probe  
 CC assays for the detection, identification, or quantification of a  
 CC prokaryotic or eukaryotic organism. cDNA transcripts of tmRNA molecules  
 CC may also be used as probes for in vitro or in situ nucleic acid  
 CC hybridisation assays. A fragment of the ssrA gene or a tmRNA molecule  
 CC corresponding to a region of high homology from the 5' end or the 3'  
 CC end of the DNA molecule can be used as a universal target region in  
 CC a nucleic acid probe assay, while a fragment of the ssrA gene or a  
 CC tmRNA molecule corresponding to a region of low homology can be used  
 CC as a target region to distinguish between species and as a target  
 CC region for the generation of genus-specific probes. These regions  
 CC may also be used as the basis for amplification primer design. The  
 CC target regions may be used as the basis of an assay for distinguishing  
 CC between living and dead prokaryotic or eukaryotic organisms, and in  
 CC a multiple probe format for broad scale detection and/or identification  
 CC of prokaryotic or eukaryotic organisms. An ssrA gene probe or a  
 CC tmRNA transcript probe can be linked to a microarray gene chip system  
 CC for the broad scale high throughput detection and identification of  
 CC prokaryotic or eukaryotic organisms. A fragment of the ssrA gene or  
 CC the tmRNA transcript can be used in an assay to obtain a DNA profile  
 CC of a prokaryotic organism and distinguish between strains of the same  
 CC species. The ssrA gene, the tmRNA transcript, DNA complementary to  
 CC an ssrA gene or tmRNA, or a fragment thereof can be used to design an  
 CC agent directed against infectious prokaryotic or eukaryotic organisms  
 CC for therapeutic purposes, and target regions may be used to monitor the  
 CC efficacy of drug therapies against infectious agents. Target regions may  
 CC also be used to monitor the viability and level of probiotic organisms  
 CC in the gastrointestinal tract. The methods and nucleic acids and  
 CC compositions of the invention have applications in medicine, and also  
 CC in industry (e.g., for assessing bacterial contamination of a foodstuff  
 CC or an environmental sample). Sequences AAF15338-F15442 represent ssrA  
 CC genes, or fragments thereof, from a wide variety of organisms.  
 XX Sequence 353 BP; 94 A; 94 C; 99 G; 66 T; 0 other;  
 SQ

Query Match 73.6%; Score 16.2; DB 22; Length 353;  
 Best Local Similarity 85.7%; Pred. No. 57;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagacag 21  
 ||||| ||||| ||||| |||||  
 Db 166 aacgagtgcagctagacag 186

RESULT 7  
 AAF15493  
 ID AAF15493 standard; RNA; 353 BP.  
 AC AAF15493;  
 XX  
 XX 09-MAR-2001 (first entry)  
 XX  
 DE Pseudomonas aeruginosa tmRNA, SEQ ID NO:102.

ssrA gene; tmRNA; bacterium; chloroplast; diatom; detection;  
 detection; identification; quantification; characterisation;  
 nucleic acid array; DNA chip; drug design; treatment monitoring;  
 contamination; ss.  
 Pseudomonas aeruginosa.  
 WO200070086-A1.  
 23-NOV-2000.  
 15-MAY-2000; 2000WO-TE00066.  
 14-MAY-1999; 99WO-TE00043.  
 (IRBI-) ENTERPRISE IRELAND T/A BIORESEARCH IRELA.  
 (UYNA-) UNIV NAT IRELAND GALWAY.  
 Barry TG, Smith TJ;  
 WPI; 2001-025025/03.  
 Use of ssrA gene, tmRNA, or fragments of them, as target regions in  
 probe assays for detection of prokaryotic or eukaryotic organisms, and  
 for determination of species -  
 Disclosure; Page 55; 221pp; English.  
 The invention relates to the ssrA gene or tmRNA, an RNA transcript  
 of the ssrA gene, or fragments thereof as target regions in a nucleic  
 acid probe assay for the detection and identification of prokaryotic  
 and/or eukaryotic organisms. The invention also relates to 38 novel ssrA  
 sequences and their tmRNA transcripts (claimed), and to pan-bacterial,  
 genus- and species-specific ssrA gene/tmRNA-directed probes and PCR  
 primers (claimed). tmRNA is a stable, high copy number RNA which is  
 found in all bacteria and is also found in chloroplasts and diatoms. It  
 has a dual function both as a tRNA and as an mRNA and is involved in  
 rescuing truncated mRNAs which have lost stop codons. ssrA genes and  
 tmRNA sequences can be used as target regions in nucleic acid probe  
 assays for the detection, identification, or quantification of a  
 prokaryotic or eukaryotic organism. cDNA transcripts of tmRNA molecules  
 may also be used as probes for in vitro or in situ nucleic acid  
 hybridisation assays. A fragment of the ssrA gene or a tmRNA molecule  
 corresponding to a region of high homology from the 5' end or the 3'  
 end of the DNA molecule can be used as a universal target region in  
 a nucleic acid probe assay, while a fragment of the ssrA gene or a  
 tmRNA molecule corresponding to a region of low homology can be used  
 as a target region to distinguish between species and as a target  
 region for the generation of genus-specific probes. These regions  
 may also be used as the basis for amplification primer design. The  
 target regions may be used as the basis of an assay for distinguishing  
 between living and dead prokaryotic or eukaryotic organisms, and in  
 a multiple probe format for broad scale detection and/or identification  
 of prokaryotic or eukaryotic organisms. An ssrA gene probe or a  
 tmRNA transcript probe can be linked to a microarray gene chip system  
 for the broad scale high throughput detection and identification of  
 prokaryotic or eukaryotic organisms. A fragment of the ssrA gene or  
 the tmRNA transcript can be used in an assay to obtain a DNA profile  
 of a prokaryotic organism and distinguish between strains of the same  
 species. The ssrA gene, the tmRNA transcript, DNA complementary to  
 an ssrA gene or tmRNA, or a fragment thereof can be used to design an  
 agent directed against infectious prokaryotic or eukaryotic organisms  
 for therapeutic purposes, and target regions may be used to monitor the  
 efficacy of drug therapies against infectious agents. Target regions may  
 also be used to monitor the viability and level of probiotic organisms  
 in the gastrointestinal tract. The methods and nucleic acids and  
 compositions of the invention have applications in medicine, and also  
 in industry (e.g., for assessing bacterial contamination of a foodstuff  
 or an environmental sample). Sequences AAF15338-F15442 represent ssrA  
 genes, or fragments thereof, from a wide variety of organisms.  
 Sequence 353 BP; 94 A; 94 C; 99 G; 66 T; 0 other;

```

Query Match      73.6%; Score 16.2; DB 22; Length 353;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aacgagtgctcagctagaccag 21
    ||||| :||| :||| |||
Db 166 aacgacugacagauagaacag 186

RESULT 8
AAH05632
ID AAH05632 standard; cDNA; 716 BP.
XX AC AAH05632;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (5'-primer) SEQ ID NO:2467.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX FN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 1; SEQ ID 2467; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesising 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesising polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.

Query Match      71.8%; Score 15.8; DB 22; Length 716;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gagtgctcagctagaccagc 22
    ||||| ||||| |||||
Db 122 gagtggcagctggaccagc 140

RESULT 9
AAH07552
ID AAH07552 standard; cDNA; 850 BP.
XX AC AAH07552;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (5'-primer) SEQ ID NO:4387.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX FN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 1; SEQ ID 4387; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesising 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesising polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
```

CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

SQ Sequence 850 BP; 196 A; 227 C; 258 G; 160 T; 9 other;

Query Match 71.8%; Score 15.8; DB 22; Length 850;  
 Best Local Similarity 89.5%; Pred. No. 1e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gaggtcagctagaccagc 22  
 ||||| ||||| |||||  
 Db 126 gagtggcagctggaccagc 144

RESULT 10  
 AAH14119  
 ID AAH14119 standard; cDNA; 1790 BP.

AC AAH14119;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:11310.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 11310; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

SQ Sequence 1790 BP; 398 A; 456 C; 504 G; 432 T; 0 other;

Query Match 71.8%; Score 15.8; DB 22; Length 1790;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gagtgcagctagaccagc 22  
 ||||| ||||| |||||  
 Db 122 gagtggcagctggaccagc 140

RESULT 11  
 AAH16554  
 ID AAH16554 standard; cDNA; 1929 BP.

XX AC AAH16554;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:15615.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 15615; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX SQ Sequence 1929 BP; 438 A; 489 C; 536 G; 466 T; 0 other;

Query Match 71.8%; Score 15.8; DB 22; Length 1929;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gagtgcagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 126 gagtgcagctgagaccagc 144

RESULT 12  
 AAD05091  
 ID AAD05091 standard; cDNA; 2096 BP.

XX AC AAD05091;

XX DT 17-JUL-2001 (first entry)

XX DE Human secreted protein-encoding gene 8 cDNA clone HSYAZ50, SEQ ID NO:49.

XX KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnery; gene therapy;  
 KW cell culture; chemotaxis; food additive;  
 KW binding partner identification; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 XX CDS 2..841

FT /\*tag= a  
 FT /product= "Human secreted protein"  
 FT /note= "CDS does not include start codon"

FT /partial

FT sig\_peptide 2..4

FT /\*tag= b

FT mat\_peptide 5..838

FT /\*tag= c  
 FT /product= "Mature human secreted protein"

XX WO200134768-A2.

XX PN 17-MAY-2001.

XX PD 01-NOV-2000; 2000WO-US30039.

XX PF 09-NOV-1999; 99US-0164344.

XX PR 07-APR-2000; 2000US-0195296.

XX PR 27-JUL-2000; 2000US-0221367.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Olsen HS, Komatsoulis G, Duan DR, Ebner R, Ruben SM;

XX WPI; 2001-308780/32.

XX DR P-FSDB; AAE01202.

XX Isolated nucleic acid molecule encoding a human secreted protein is

PT used in preventing, treating or ameliorating a medical condition -  
 XX Claim 1; Page 429; 474pp; English.

XX AAD05053-AAD05106 represent cDNAs corresponding to 15 human secreted  
 CC protein genes, and AAE01164-AAE01217 represent the proteins they encode.  
 CC AAE01218-AAE01226 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 15 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis or treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC allergies, neurological disorders (e.g., Alzheimer's disease,  
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,  
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
 CC disorders, and infections. The proteins can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues, to identify their  
 CC cognate ligands or binding partners, and in chemotaxis, and can be used  
 CC as a food additive or preservative to modify storage properties.  
 CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 CC secreted protein-encoding cDNA of the invention.

SQ Sequence 2096 BP; 476 A; 529 C; 584 G; 504 T; 3 other;

Query Match 71.8%; Score 15.8; DB 22; Length 2096;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gagtgcagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 149 gagtgcagctgagaccagc 167

RESULT 13

AAV69286/c  
 ID AAV69286 standard; DNA; 10708 BP.

XX AC AAV69286;

XX DT 02-FEB-1999 (first entry)

XX DE Sequence of mouse activin genetic loci.

XX KW Liver activin; beta c; beta e; cell differentiation; haematopoiesis;  
 KW erythroid; ovarian follicular maturation; hormone; neuronal survival;  
 KW spermatogenesis; bone; insulin; cardiac morphogenesis; osteoporosis;  
 KW osteomalacia; erythropoiesis; haemophilia; cystic fibrosis; immunoassay;  
 KW menstrual disorder; transgenic; modulator; ss.

XX OS Mus sp.

XX FH Key Location/Qualifiers  
 XX CDS 1147..2736

FT /\*tag= a  
 FT /product= activin beta c"

FT /note= "contains introns"

FT exon 1147..1462

FT /\*tag= b  
 FT /number= 1  
 FT /note= "activin beta c exon 1"

FT intron 1463..1994  
 FT /\*tag= c  
 FT /number= 1  
 FT /note= "activin beta c partial intron (12-kb)"  
 FT exon 1995..2733  
 FT /\*tag= d  
 FT /number= 2  
 FT 2737..8248  
 FT /tag= e  
 FT /note= "activin beta e promoter region"  
 FT CDS 8249..9536  
 FT /\*tag= f  
 FT /product= "activin beta e"  
 FT /note= "contains introns"  
 FT exon 8249..8546  
 FT /\*tag= g  
 FT /number= 1  
 FT /note= "activin beta e exon 1"  
 FT 8547..8782  
 FT /\*tag= h  
 FT /number= 1  
 FT /note= "activin beta e intron"  
 FT exon 8783..9533  
 FT /\*tag= i  
 FT /number= 2  
 FT /note= "activin beta e exon 2"  
 FT XX  
 PN W09822492-A1.  
 XX 28-MAY-1998.  
 XX 20-NOV-1997; 97WO-US20882.  
 XX 20-NOV-1996; 96US-0752919.  
 XX (UNMI ) UNIV MICHIGAN.  
 XX Bonadio J, Fang J;  
 XX WPI; 1998-312408/27.  
 XX P-PSDB; AAW60617, AAW60618.  
 XX New isolated nucleic acid encoding sub-units of liver activin -  
 XX useful for regulating growth and differentiation of cells, e.g. for  
 XX treating liver, bone and haematopoietic disorders  
 XX Disclosure; Fig 4D-G; 141pp; English.  
 XX This represents the sequence of mouse activin genetic loci. The invention  
 XX relates to murine beta c and beta e polypeptides and the genes encoding  
 XX them. Disorders of cell growth or differentiation (or susceptibility to  
 XX them) are diagnosed by measuring liver activin gene activity or by  
 XX detecting a mutation in the liver activin gene. Disorders of  
 XX haematopoiesis, erythroid differentiation, ovarian follicular maturation,  
 XX hormone secretion, neuronal survival, spermatogenesis, bone formation,  
 XX insulin secretion or cardiac morphogenesis are some conditions that can  
 XX be diagnosed using the liver activin. Cell growth and differentiation can  
 XX be stimulated by treatment with an liver activin compound or agent that  
 XX upregulates the compound's expression. Antagonists can be used to treat  
 XX liver diseases while agonists can be used to increase growth and  
 XX regeneration of liver tissue. The liver activin compound may also induce  
 XX bone growth (e.g. for treating osteoporosis or osteomalacia) or  
 XX haematopoiesis, particularly erythropoiesis, e.g. for treating  
 XX useful in immunoassays, to generate anti-idiotypic antibodies (which bind  
 XX to liver activin receptors) and to inhibit liver activin. Also,  
 XX transgenic animals containing liver activin gene can be used to produce  
 XX the liver activin (in correctly processed and modified forms) proteins,  
 XX or the transgenic animals, are useful for screening for liver activin  
 XX modulators.  
 XX This cDNA encodes a murine liver activin beta c polypeptide. Sequences  
 XX derived from beta c cDNA clone is used for screening and cloning a liver  
 XX activin beta e gene. Disorders of cell growth or differentiation (or

CC susceptibility to them) are diagnosed by measuring liver activin gene  
 CC activity or by detecting a mutation in the liver activin gene. Disorders  
 CC of haematopoiesis, erythroid differentiation, ovarian follicular  
 CC maturation, hormone secretion, neuronal survival, spermatogenesis, bone  
 CC formation, insulin secretion or cardiac morphogenesis are some conditions  
 CC that can be diagnosed using the liver activin. Cell growth and  
 CC differentiation can be stimulated by treatment with an liver activin  
 CC compound or agent that upregulates the compound's expression. Antagonists  
 CC can be used to treat liver diseases while agonists can be used to  
 CC increase growth and regeneration of liver tissue. The liver activin  
 CC compound may also induce bone growth (e.g. for treating osteoporosis or  
 CC osteomalacia) or haematopoiesis, particularly erythropoiesis, e.g. for  
 CC treating haemophilia, cystic fibrosis or menstrual disorders. Antibodies  
 CC are useful in immunoassays, to generate anti-idiotypic antibodies (which  
 CC bind to liver activin receptors) and to inhibit liver activin. Also,  
 CC transgenic animals containing liver activin gene can be used to produce  
 CC the liver activin (in correctly processed and modified forms) proteins,  
 CC or the transgenic animals, are useful for screening for liver activin  
 CC modulators.  
 CC XX

SQ Sequence 10708 BP; 2651 A; 2844 C; 2607 G; 2546 T; 60 other;

Query Match 71.8%; Score 15.8; DB 19; Length 10708;  
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gagtgctagctagaccagc 22  
 ||||| ||| |||||  
 Db 1949 GAGTGTCTGCTGCACCAGC 1931

## RESULT 14

AA152779  
 ID AAI52779 standard; DNA; 163 BP.

XX AAI52779;

DT 17-OCT-2001 (first entry)

DE Probe #21465 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 23-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 21465; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).

**Qy**      1   aacgagtgtcagctagaccgc 22  
           ||||| |||||||  
**Db**      257 aaqgaqtcttgcctagaccgc 278

-----



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:52 ; Search time 301.6 Seconds  
(without alignments)  
16.520 Million cell updates/sec

Title: US-09-698-903B-13

Perfect score: 22

Sequence: 1 aacgagtgcagtagaccagc 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/1na/PTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/1na/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	3113	1	Sequence 20, Appl
2	22	100.0	3113	1	Sequence 2, Appl
C 3	15.4	70.0	1330	3	Sequence 5, Appl
C 4	15.4	70.0	2600	1	Sequence 1, Appl
C 5	15.2	69.1	350	2	Sequence 16, Appl
C 6	15.2	69.1	350	2	Sequence 16, Appl
C 7	15.2	69.1	350	3	Sequence 16, Appl
C 8	15.2	69.1	819	4	Sequence 2, Appl
C 9	15.2	69.1	1550	2	Sequence 17, Appl
C 10	15.2	69.1	1550	2	Sequence 17, Appl
C 11	15.2	69.1	1550	2	Sequence 2, Appl
C 12	15.2	69.1	1550	3	Sequence 17, Appl
C 13	15.2	69.1	1550	3	Sequence 3, Appl
C 14	14.6	66.4	2028	2	Sequence 73, Appl
C 15	14.6	66.4	2028	2	Sequence 73, Appl
C 16	14.6	66.4	2915	2	Sequence 1, Appl
C 17	14.6	66.4	3719	1	Sequence 10, Appl
C 18	14.6	66.4	3719	1	Sequence 10, Appl
C 19	14.6	66.4	3719	1	Sequence 10, Appl
C 20	14.6	66.4	3719	1	Sequence 10, Appl
C 21	14.6	66.4	3719	2	Sequence 10, Appl
C 22	14.6	66.4	50341	1	Sequence 1, Appl
C 23	14.6	66.4	50341	2	Sequence 1, Appl
C 24	14.6	66.4	52297	4	Sequence 1, Appl
C 25	14.6	66.4	52297	4	Sequence 1, Appl
C 26	14.6	66.4	4411529	4	Sequence 1, Appl
C 27	14.4	65.5	1458	5	Sequence 1, Appl

c 28	14.4	65.5	1461	1	US-07-596-867C-10	Sequence 10, Appl
c 29	14.4	65.5	1506	3	US-08-445-463B-1	Sequence 1, Appl
c 30	14.4	65.5	1506	3	US-08-445-464C-1	Sequence 1, Appl
c 31	14.4	65.5	1963	5	PCT-US91-07715A-1	Sequence 1, Appl
c 32	14.4	65.5	1970	1	US-07-596-867C-1	Sequence 1, Appl
c 33	14.2	64.5	192	4	US-09-060-756-634	Sequence 634, App
c 34	14.2	64.5	215	4	US-09-060-756-194	Sequence 194, App
c 35	14.2	64.5	230	4	US-09-060-756-422	Sequence 422, App
c 36	14.2	64.5	579	2	US-08-611-757-50	Sequence 50, Appl
c 37	14.2	64.5	579	5	PCT-US95-05980-50	Sequence 50, Appl
c 38	14.2	64.5	1679	6	5196194-12	Patent No. 5196194
c 39	14.2	64.5	2389	2	US-08-372-652-8	Sequence 8, Appl
c 40	14.2	64.5	2389	5	PCT-US95-16311-8	Sequence 8, Appl
c 41	14.2	64.5	2940	3	US-09-195-868-11	Sequence 11, Appl
c 42	14.2	64.5	3262	2	US-08-678-039A-41	Sequence 41, Appl
c 43	14.2	64.5	3691	3	US-09-195-868-12	Sequence 12, Appl
c 44	14.2	64.5	4147	3	US-08-560-005-1	Sequence 1, Appl
c 45	14.2	64.5	4147	4	US-09-418-540-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-146-422-20  
; Sequence 20, Application US/08146422  
; Patent No. 5543576  
; GENERAL INFORMATION:  
; APPLICANT: VAN OOLJEN, ALBERT J. J.  
; APPLICANT: RIETVELD, KRIJN  
; APPLICANT: HOEKEMA, ANDREAS  
; APPLICANT: PEN, JAN  
; APPLICANT: SIJMONS, PETER C.  
; APPLICANT: VERWOERD, TEUNIS C.  
; APPLICANT: QUAX, WILHEMUS J.  
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/146.422  
; FILING DATE: 02-NOV-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KENNEDY, BILL  
; REGISTRATION NUMBER: 33,407  
; REFERENCE/DOCKET NUMBER: 44615-20011.23  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3113 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-146-422-20

Query Match 100.0%; Score 22; DB 1; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.042;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgctcagctagaccagc 22  
Db 790 AACGAGTGCTCAGCTAGACCAGC 811

## RESULT 2

US-08-626-554-2  
; Sequence 2, Application US/08626554  
; Patent No. 5714474  
; GENERAL INFORMATION:  
; APPLICANT: VAN OOIJEN, ALBERT J.J.  
; APPLICANT: RIETVELD, KRIJN  
; APPLICANT: HOEKEMA, ANDREAS  
; APPLICANT: PEN, JAN  
; APPLICANT: SIJMONS, PETER C.  
; APPLICANT: VERWOERD, TEUNIS C.  
; APPLICANT: QUAX, WILHEMUS J.  
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
; TITLE OF INVENTION: USE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 PENNSYLVANIA AVENUE NW  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/626,554  
; FILING DATE: 02-APR-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 26192-20011.10  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030 MRSNFOERSWSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3113 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-626-554-2

Query Match 100.0%; Score 22; DB 1; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.042;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgctcagctagaccagc 22  
Db 790 AACGAGTGCTCAGCTAGACCAGC 811

## RESULT 3

US-08-147-592A-5/c  
; Sequence 5, Application US/08147592A  
; Patent No. 6096513  
; GENERAL INFORMATION:  
; APPLICANT: Bell, Graeme I  
; APPLICANT: Relsine, Terry  
; APPLICANT: Yasuda, Kazuki  
; TITLE OF INVENTION: Opioid Receptor Genes,

; TITLE OF INVENTION: Compositions and Methods  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: United States of America  
; ZIP: 72210

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/147,592A  
; FILING DATE: 05-NOV-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/100,694  
; FILING DATE: 30-JUL-1993  
; CLASSIFICATION: 435  
; APPLICATION NUMBER: 08/066,296  
; FILING DATE: 20-MAY-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wilson, Mark B.  
; REGISTRATION NUMBER: 37,259  
; REFERENCE/DOCKET NUMBER: ARCD:105  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; TELEX: N/A  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1330 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 161..1261  
US-08-147-592A-5

Query Match 70.0%; Score 15.4; DB 3; Length 1330;  
Best Local Similarity 94.1%; Pred. No. 63;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 agtgcagctagaccag 21  
Db 462 AGTGCAGCAGACCAG 446

## RESULT 4

US-08-147-949A-1/c  
; Sequence 1, Application US/08147949A  
; Patent No. 5747279  
; GENERAL INFORMATION:  
; APPLICANT: Pasternak, Gavrill W.  
; APPLICANT: Pan, Ying-Xian  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: KAPPA3 OPIOID RECEPTORS, RECEPTORS  
; TITLE OF INVENTION: ENCODED THEREBY, AND USES THEREOF  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

RESULT 5  
US-08-466-337A-16/c  
; Sequence 16, Application US/08466337A  
; Patent No. 5830756  
; GENERAL INFORMATION:  
; APPLICANT: Haskill, John S.  
; APPLICANT: Baldwin Jr., Albert S.  
; APPLICANT: Ralph, Peter  
; TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional  
; TITLE OF INVENTION: Activator and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower/ 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,337A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pochopien, Donald J.  
; REGISTRATION NUMBER: 32,167  
; REFERENCE/DOCKET NUMBER: 0899.008/33518  
; TELECOMMUNICATION INFORMATION:

Qy 3 cgagtgcagctagaccgc 22

4

TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 95...1045  
US-08-466-337A-17

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgctcagctagaccagc 22  
| ||||| |||||  
Db 917 CTAGTGTGAGTGGCCAGC 898

RESULT 10  
US-08-475-359-17/c  
; Sequence 17, Application US/08475359  
; Patent No. 5846714  
; GENERAL INFORMATION:  
; APPLICANT: Haskill, John S.  
; APPLICANT: Baldwin Jr., Albert S.  
; APPLICANT: Ralph, Peter  
; TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional  
; TITLE OF INVENTION: Activator and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower/ 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60608-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,359  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pochopien, Donald J.  
; REGISTRATION NUMBER: 32,167  
; REFERENCE/DOCKET NUMBER: 0899,004,33514  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1550 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 95...1045  
US-08-475-359-17

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgctcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGTGGCCAGC 898

RESULT 11  
US-08-802-322-2/c  
; Sequence 2, Application US/08802322  
; Patent No. 5932425  
; GENERAL INFORMATION:  
; APPLICANT: Alkalay, Irit  
; APPLICANT: Ben-Neriah, Yinon  
; APPLICANT: Ciechanover, Aaron  
; APPLICANT: Manning, Anthony  
; APPLICANT: Mercurio, Frank  
; APPLICANT: Yaron, Avraham  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING  
; TITLE OF INVENTION: CELLULAR NF-(B ACTIVATION  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/802,322  
; FILING DATE: 18-FEB-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 860098.418  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1550 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-802-322-2

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgctcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGTGGCCAGC 898

RESULT 12  
US-08-465-887A-17/c  
; Sequence 17, Application US/08465887A  
; Patent No. 6001582  
; GENERAL INFORMATION:  
; APPLICANT: Haskill, John S.  
; APPLICANT: Baldwin Jr., Albert S.  
; APPLICANT: Ralph, Peter  
; TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional

;; TITLE OF INVENTION: Activator and Uses Thereof  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
;; STREET: 6300 Sears Tower/ 233 South Wacker Drive  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: United States  
;; ZIP: 60606-6402  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/465,887A  
;; FILING DATE: 06-JUN-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fochopien, Donald J.  
;; REGISTRATION NUMBER: 32,167  
;; REFERENCE/DOCKET NUMBER: 0899.006/33516  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312/474-6300  
;; TELEFAX: 312/474-0448  
;; TELEX: 25-3856  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1550 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 95..1045  
US-08-465-887A-17

Query Match 69.1%; Score 15.2; DB 3; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgtcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGCTGCCGAGC 898

RESULT 13  
US-08-895-601-3/c  
; Sequence 3, Application US/08895601  
; Patent No. 6060262  
; GENERAL INFORMATION:  
; APPLICANT: Beer-Romero, Peggy  
; APPLICANT: Strack, Peter J.  
; APPLICANT: Glass, Susan J.  
; APPLICANT: Rolfe, Mark  
; TITLE OF INVENTION: REGULATION OF KAPPA B (1kb) DEGRADATION,  
; AND METHODS AND REAGENTS RELATED THERETO  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/895,601  
;; FILING DATE: 16-JUL-1997  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Vincent, Matthew P.  
;; REGISTRATION NUMBER: 36,709  
;; REFERENCE/DOCKET NUMBER: MIV-096.01  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-832-1000  
;; TELEFAX: 617-832-7000  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1550 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 95..1045  
US-08-895-601-3

Query Match 69.1%; Score 15.2; DB 3; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgtcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGCTGCCGAGC 898

RESULT 14  
US-08-933-750C-73/c  
; Sequence 73, Application US/08933750C  
; Patent No. 5932442  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,750C  
; FILING DATE: September 23, 1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0356 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166

; TELEX:  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2028 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: UTRSNOT05  
; CLONE: 1568361  
US-08-933-750C-73

Query Match 66.4%; Score 14.6; DB 2; Length 2028;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aacgagtgctcagctagaccag 21  
| | | | | | | | | | | | | | |  
Db 1486 ATCCAGTGTGTCAGCAACACCAG 1466

RESULT 15  
US-09-234-613-73/c  
; Sequence 73, Application US/09234613  
; Patent No. 6132973  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/234,613  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,750  
; FILING DATE: September 23, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0356 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2028 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: UTRSNOT05  
; CLONE: 1568361

US-09-234-613-73

Query Match 66.4%; Score 14.6; DB 3; Length 2028;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 aacgagtgctcagctagaccag 21  
| | | | | | | | | | | | | | |  
Db 1486 ATCCAGTGTGTCAGCAACACCAG 1466

Search completed: February 25, 2002, 18:05:55  
Job time: 18603 sec

-----



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:20 ; Search time 8261.74 Seconds  
(without alignments)  
28.615 Million cell updates/sec

Title: US-09-698-903B-13

Perfect score: 22

Sequence: 1 aacgagtgtagcagaccagc 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	85.5	152	10	BE520593
2	18.8	85.5	174	10	BE523925
3	18.8	85.5	182	10	BE524190
4	18.8	85.5	206	10	BE525280
5	18.8	85.5	212	10	BE523695
6	18.8	85.5	223	10	BE524251
7	18.8	85.5	225	10	BE523875
8	18.8	85.5	231	10	BE522306
9	18.8	85.5	232	10	BE520764
10	18.8	85.5	232	10	BE525569
11	18.8	85.5	238	10	BE525596
12	18.8	85.5	239	10	BE523203

13	18.8	85.5	243	10	BE525564
14	18.8	85.5	244	10	BE525454
15	18.8	85.5	244	10	BE525533
16	18.8	85.5	245	10	BE523453
17	18.8	85.5	247	10	BE523445
18	18.8	85.5	249	10	BE525542
19	18.8	85.5	252	10	BE524837
20	18.8	85.5	253	10	BE525432
21	18.8	85.5	253	10	BE525500
22	18.8	85.5	254	10	BE522200
23	18.8	85.5	254	10	BE524936
24	18.8	85.5	255	10	BE525465
25	18.8	85.5	261	10	BE522627
26	18.8	85.5	272	10	BE523398
27	18.8	85.5	272	10	BE524178
28	18.8	85.5	280	10	BE520917
29	18.8	85.5	283	10	BE520827
30	18.8	85.5	287	10	BE522480
31	18.8	85.5	291	10	BE521713
32	18.8	85.5	298	10	BE523183
33	18.8	85.5	300	10	BE525377
34	18.8	85.5	303	10	BE525401
35	18.8	85.5	305	10	BE520824
36	18.8	85.5	305	10	BE521646
37	18.8	85.5	305	10	BE524379
38	18.8	85.5	306	10	BE522847
39	18.8	85.5	306	10	BE523254
40	18.8	85.5	307	10	BE523644
41	18.8	85.5	308	10	BE524918
42	18.8	85.5	310	10	BE520427
43	18.8	85.5	311	10	BE520918
44	18.8	85.5	312	10	BE520556
45	18.8	85.5	312	10	BE522759

#### ALIGNMENTS

RESULT 1

BE520593

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BE520593 152 bp mRNA EST 19-MAR-2001  
M13F7STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
clone M13F7 5', mRNA sequence.  
BE520593  
BE520593.1 GI:9778571  
EST.  
Arabidopsis thaliana  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 152)  
Whitney, J.A., Todd, J., Newman, T., Focke, N., Girke, T., Martinez de  
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Michigan State University DNA Sequencing Facility Arabidopsis  
Biological Resource Center, The Ohio State University, 309 Botany &  
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:  
6142920603 TEL: 6142929371.  
Location/Qualifiers  
1. .152  
/organism="Arabidopsis thaliana"

```

/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F7"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
34 t

BASE COUNT      29 a 47 c 42 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 152;
Best Local Similarity 90.9%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
    ||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 2
BE523925 174 bp mRNA EST 19-MAR-2001
LOCUS M43F1STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M43F11 5', mRNA sequence.
ACCESSION BE523925
VERSION BE523925.1 GI:9781903
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 174)
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlroge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .174
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F11"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
46 t

BASE COUNT      36 a 52 c 40 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 174;
Best Local Similarity 90.9%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 aacgagtgcagctagaccagc 22
    ||||| ||||| ||||| |||||
Db 104 AACGAGTCCAGCTCGACCAGC 125

RESULT 3
BE524190 182 bp mRNA EST 19-MAR-2001
LOCUS M47A2STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M47A2 5', mRNA sequence.
ACCESSION BE524190
VERSION BE524190.1 GI:9782168
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 182)
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlroge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .182
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M47A2"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
42 t

BASE COUNT      33 a 58 c 49 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 182;
Best Local Similarity 90.9%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
    ||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 4
BE525280 206 bp mRNA EST 19-MAR-2001
LOCUS M61A10STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M61A10 5', mRNA sequence.
ACCESSION BE525280
VERSION BE525280.1 GI:9783258
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis; 1 (bases 1 to 206)  
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J., and Benning, C.  
A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel.: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

#### FEATURES

source

1..206  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="M61A10"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: ECORI; Site\_2: XhoII"  
44 a 58 c 54 g 50 t

BASE COUNT  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 206;  
Best Local Similarity 90.9%; Pred. No. 82;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
||||||| ||||| |||||

Db 121 AACGAGTCCAGCTCGACCAGC 142

RESULT 5

BE523695 212 bp mRNA EST 19-MAR-2001  
LOCUS M40E7STM Arabidopsis developing seed Arabidopsis thaliana CDNA  
DEFINITION BE523695  
ACCESSION BE523695  
VERSION BE523695.1 GI:9781673  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis; 1 (bases 1 to 212)  
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J., and Benning, C.

A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel.: 517 355 1609

FEATURES  
source

1..223  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="M47G12"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"

Fax: 517 353 9334  
Email: benning@msu.edu

Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

#### FEATURES

source

1..212  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="M40E7"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: ECORI; Site\_2: XhoII"  
42 a 73 c 45 g 52 t

BASE COUNT  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 212;  
Best Local Similarity 90.9%; Pred. No. 83;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
||||||| ||||| |||||

Db 24 AACGAGTCCAGCTCGACCAGC 45

RESULT 6

BE524251 223 bp mRNA EST 19-MAR-2001  
LOCUS M47G12STM Arabidopsis developing seed Arabidopsis thaliana CDNA  
DEFINITION BE524251  
ACCESSION BE524251  
VERSION BE524251.1 GI:9782229  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis; 1 (bases 1 to 223)  
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J., and Benning, C.

A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel.: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu

Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

FEATURES

source

1..223  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="M47G12"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"

```

/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcorI; Site_2: XhoII"
BASE COUNT      37 a 69 c 57 g 60 t
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 223;
Best Local Similarity 90.9%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22
    ||||| ||||| ||||| |||||
Db 99 AACGAGTCCAGCTCGACCAGC 120

RESULT 7
LOCUS      BE523875      225 bp      mRNA      EST      19-MAR-2001
DEFINITION M43A1STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION  BE523875
VERSION     BE523875.1 GI:9781853
KEYWORDS    EST.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE   1 (bases 1 to 225)
AUTHORS     White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE       A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL     Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE     20567808
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI 48824
            , USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Michigan State University DNA Sequencing Facility Arabidopsis
            Biological Resource Center, The Ohio State University, 309 Botany &
            Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
            6142920603 TEL: 6142929371.
            Location/Qualifiers
FEATURES             source
     1..225
     /organism="Arabidopsis thaliana"
     /strain="Columbia"
     /db_xref="taxon:3702"
     /clone="M43A1"
     /clone_lib="Arabidopsis developing seed"
     /tissue_type="seed"
     /dev_stage="5-13 days after flowering"
     /lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcorI; Site_2: XhoII"
BASE COUNT      48 a 68 c 52 g 56 t
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 225;
Best Local Similarity 90.9%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22
    ||||| ||||| ||||| |||||
Db 118 AACGAGTCCAGCTCGACCAGC 139

RESULT 8

```

```

BE522306      231 bp      mRNA      EST      19-MAR-2001
LOCUS      M25A1STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION  clone M25A1 5', mRNA sequence.
ACCESSION  BE522306
VERSION     BE522306.1 GI:9780284
KEYWORDS    EST.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE   1 (bases 1 to 231)
AUTHORS     White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE       A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL     Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE     20567808
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI 48824
            , USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Michigan State University DNA Sequencing Facility Arabidopsis
            Biological Resource Center, The Ohio State University, 309 Botany &
            Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
            6142920603 TEL: 6142929371.
            Location/Qualifiers
FEATURES             source
     1..231
     /organism="Arabidopsis thaliana"
     /strain="Columbia"
     /db_xref="taxon:3702"
     /clone="M25A1"
     /clone_lib="Arabidopsis developing seed"
     /tissue_type="seed"
     /dev_stage="5-13 days after flowering"
     /lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcorI; Site_2: XhoII"
BASE COUNT      45 a 72 c 57 g 57 t
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 231;
Best Local Similarity 90.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22
    ||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 9
LOCUS      BE520764      232 bp      mRNA      EST      19-MAR-2001
DEFINITION  M14H9STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION  BE520764
VERSION     BE520764.1 GI:9778742
KEYWORDS    EST.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE   1 (bases 1 to 232)
AUTHORS     White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE       A new set of Arabidopsis expressed sequence tags from developing

```

JOURNAL MEDLINE COMMENT

seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)

Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabisopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.

FEATURES

source

1. .232  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014455R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
Location/Qualifiers

BASE COUNT 44 a 73 c 57 g 58 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 232;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22  
||||||| ||||| ||||| |||||

Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 10

BE525569

LOCUS

DEFINITION M62M13STM Arabidopsis developing seed Arabidopsis thaliana cDNA

ACCESSION BE525569

VERSION BE525569.1 GI:9783470

KEYWORDS EST.

SOURCE

ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 232)  
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Javorski,J.G., Ohlrogge,J. and Benning,C.  
A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)

JOURNAL MEDLINE COMMENT

Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabisopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.

FEATURES

source

1. .232  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014455R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
Location/Qualifiers

BASE COUNT 44 a 73 c 57 g 58 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 232;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22  
||||||| ||||| ||||| |||||

Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 10

BE525569

LOCUS

DEFINITION M62M13STM Arabidopsis developing seed Arabidopsis thaliana cDNA

ACCESSION BE525569

VERSION BE525569.1 GI:9783470

KEYWORDS EST.

SOURCE

ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 232)  
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Javorski,J.G., Ohlrogge,J. and Benning,C.  
A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)

JOURNAL MEDLINE COMMENT

Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabisopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.

FEATURES

source

1. .232  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014455R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
Location/Qualifiers

BASE COUNT 45 a 70 c 63 g 58 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 238;

source

1. .232  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014455R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
Location/Qualifiers

BASE COUNT 41 a 72 c 56 g 63 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 232;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22  
||||||| ||||| ||||| |||||

Db 90 AACGAGTCCAGCTCGACCAGC 111

RESULT 11

BE525596

LOCUS

DEFINITION M62C23STM Arabidopsis developing seed Arabidopsis thaliana cDNA

ACCESSION BE525596

VERSION BE525596.1 GI:9783497

KEYWORDS EST.

SOURCE

ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 238)  
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Javorski,J.G., Ohlrogge,J. and Benning,C.  
A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)

JOURNAL MEDLINE COMMENT

Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabisopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.

FEATURES

source

1. .238  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014490R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
Location/Qualifiers

BASE COUNT 45 a 70 c 63 g 58 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 238;

Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22  
||||| ||||| ||||| |||||  
Db 24 AACGAGTCCAGCTCGACCAGC 45

RESULT 12  
BE523203  
LOCUS M35C7STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
DEFINITION clone M35C7 5', mRNA sequence.  
ACCESSION BE523203  
VERSION BE523203.1 GI:9781277  
KEYWORDS EST.

SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 239)  
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.  
TITLE A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)  
MEDLINE 20567808  
COMMENT Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142929371.

FEATURES  
source  
1..239  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="M35C7"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
BASE COUNT 49 a 74 c 57 g 59 t  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 239;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 aacgagtgcagctagaccagc 22  
||||| ||||| ||||| |||||  
Db 113 AACGAGTCCAGCTCGACCAGC 134

RESULT 13  
BE525564  
LOCUS M62C13STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
DEFINITION clone 600014450R1 5', mRNA sequence.  
ACCESSION BE525564  
VERSION BE525564.1 GI:9783465  
KEYWORDS EST.

SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 244)  
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.  
TITLE A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)  
MEDLINE 20567808  
COMMENT Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:  
6142929371.

SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
COMMENT

FEATURES  
source

1..243  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014450R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
BASE COUNT 49 a 69 c 65 g 59 t  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 243;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 aacgagtgcagctagaccagc 22  
||||| ||||| ||||| |||||  
Db 21 AACGAGTCCAGCTCGACCAGC 42

RESULT 14  
BE525454  
LOCUS M62G03STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
DEFINITION clone 600014412R1 5', mRNA sequence.  
ACCESSION BE525454  
VERSION BE525454.1 GI:9783432  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 244)  
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.  
TITLE A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)  
MEDLINE 20567808  
COMMENT Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142929371.

, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.

## FEATURES

source

1. .244  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="600014412R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"

BASE COUNT 46 a 74 c 66 g 57 t 1 others  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 244;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
||||||| ||||| |||||  
Db 44 AACGAGTCCAGCTCGACCAGC 65

## RESULT 15

BE525533

LOCUS

BE525533 244 bp mRNA EST 19-MAR-2001  
DEFINITION M62K03STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
clone 60001441R1 5', mRNA sequence.

ACCESSION

BE525533

VERSION

BE525533.1

GI:9783434

KEYWORDS

EST

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 244)

White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de

Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.

A new set of Arabidopsis expressed sequence tags from developing

seeds. The metabolic pathway from carbohydrates to seed oil

Plant Physiol. 124 (4), 1582-1594 (2000)

20567808

Contact: Benning, C

Dept. of Biochemistry &amp; Molecular Biology

Michigan State University

224 Biochemistry, Michigan State University, East Lansing, MI 48824

, USA

Tel: 517 355 1609

Fax: 517 353 9334

Email: benning@msu.edu

Clones were originally prepared at Michigan State University.

Arabidopsis Biological Resource Center, The Ohio State University,

309 Botany &amp; Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210

USA, FAX: 6142920603 TEL: 6142929371.

Location/Qualifiers

1. .244

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone\_lib="60001441R1"

/tissue\_type="seed"

/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
43 a 74 c 60 g 67 t

BASE COUNT  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 244;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
||||||| ||||| |||||  
Db 91 AACGAGTCCAGCTCGACCAGC 112

Search completed: February 25, 2002, 17:21:21  
Job time: 16174 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:03:03 ; Search time 2331.3 Seconds  
(without alignments)  
155.680 Million cell updates/sec

Title: US-09-698-903B-14

Perfect score: 22  
Sequence: 1 ccgagtcgtgcaacatcgacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo\_hum.\*
- 31: em.htgo\_inv.\*
- 32: em.htgo\_rod.\*
- 33: em.htg\_hum.\*
- 34: em.htg\_inv.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	6	AX127761 Sequence
2	22	100.0	22	6	AX172482 Sequence
3	22	100.0	3113	6	I24540 Sequence 20
4	22	100.0	3113	6	I83673 Sequence 2
5	22	100.0	3113	8	X14555 Brassica na
6	22	100.0	3198	8	X59294 B.napus Bnc
7	17.4	79.1	143830	33	AC021599 Homo sapi
8	17.4	79.1	151210	2	AC092613 Homo sapi
9	17.4	79.1	154016	9	AC005047 Homo sapi
10	17.2	78.2	287	4	AF142615 Equus cab
11	17.2	78.2	587	6	I56095 Sequence 3
12	17.2	78.2	646	1	AF190914 Escherich
13	17.2	78.2	1174	6	I56097 Sequence 5
14	17.2	78.2	1188	6	I56101 Sequence 9
15	17.2	78.2	1196	6	I56099 Sequence 7
16	17.2	78.2	1264	1	SSIS600
17	17.2	78.2	1433	1	SF097492
18	17.2	78.2	3294	1	SHFSHTA
19	17.2	78.2	3715	1	AF081284 Escherich
20	17.2	78.2	3719	1	SF081136
21	17.2	78.2	5430	1	SF082621
22	17.2	78.2	6014	1	SD271153
23	17.2	78.2	7150	1	AF177050
24	17.2	78.2	10029	1	AE007213
25	17.2	78.2	10040	1	AE005271
26	17.2	78.2	10605	1	AF139596
27	17.2	78.2	11058	1	AE005307
28	17.2	78.2	15708	3	AF051097
29	17.2	78.2	20962	1	AF335540
30	17.2	78.2	20962	1	AF335540
31	17.2	78.2	23771	1	AF141323
32	17.2	78.2	32094	1	AF153317
33	17.2	78.2	37010	1	EC027814
34	17.2	78.2	51334	1	AF200692
35	17.2	78.2	72283	2	AC020099
36	17.2	78.2	92077	1	AF074613
37	17.2	78.2	92077	6	AX191727
38	17.2	78.2	92721	1	AB011549
39	17.2	78.2	92721	6	AX191725
40	17.2	78.2	165190	3	AC093096
41	17.2	78.2	213494	1	SFPWR100
42	17.2	78.2	213494	1	SFPWR100
43	17.2	78.2	217280	2	AC019337
44	17.2	78.2	221851	1	AF348706
45	17.2	78.2	221851	1	AF348706

ALIGNMENTS

RESULT 1  
AX127761  
LOCUS AX127761 22 bp DNA  
DEFINITION Sequence 14 from Patent WO0131042.  
ACCESSION AX127761  
VERSION AX127761.1 GI:14134408  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
artificial construct  
artificial sequence.  
1 (bases 1 to 22)  
REFERENCE  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 14 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
1..22  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer CV28"

PAT

15-MAY-2001

BASE COUNT 5 a 7 c 5 g 5 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22

Db 1 CGCAGTTCTGTGAACATCGACC 22

# RESULT 2

AX172482 AX172482 22 bp DNA PAT 03-JUL-2001  
LOCUS Sequence 43 from Patent WO0141558.  
DEFINITION AX172482  
ACCESSION AX172482  
VERSION AX172482.1 GI:14597594

KEYWORDS synthetic construct.

SOURCE synthetic construct.

ORGANISM artificial sequence.

REFERENCE 1 (bases 1 to 22)

AUTHORS de Both, G. and de Beuckeleer, M.

TITLE Hybrid winter oilseed rape and methods for producing same

JOURNAL Patent: WO 0141558-A 43 14-JUN-2001;

AVENTIS CropScience N.V. (BE)

FEATURES Location/Qualifiers

source 1..22

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer BNA06"

BASE COUNT 5 a 7 c 5 g 5 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22

Db 1 CGCAGTTCTGTGAACATCGACC 22

# RESULT 3

I24540/c I24540/c 3113 bp DNA PAT 07-OCT-1996  
LOCUS Sequence 20 from patent US 5543576.  
DEFINITION I24540  
ACCESSION I24540  
VERSION I24540.1 GI:1604410

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 3113)

AUTHORS van Ooijen, A.J.J., Rietveld, K., Hoekema, A., Pen, J., Sijmons, P.C.,

Verwoerd, T.C., and Quax, W.J.

TITLE Production of enzymes in seeds and their use

JOURNAL Patent: US 5543576-A 20 06-AUG-1996;

FEATURES Location/Qualifiers

source 1..3113

/organism="unknown"

BASE COUNT 961 a 685 c 586 g 881 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22

Db 1183 CGCAGTTCTGTGAACATCGACC 1162

# RESULT 4

I83673/c I83673 3113 bp DNA PAT 10-AUG-1998  
LOCUS Sequence 2 from patent US 5714474.  
DEFINITION I83673  
ACCESSION I83673  
VERSION I83673.1 GI:3407203

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 3113)

AUTHORS Van Ooijen, A.J.J., Rietveld, K., Hoekema, A., Pen, J.,

Sijmons, P., Christian, Verwoerd, T., Cornelis and Quax, W. Johannes.

TITLE Production of enzymes in seeds and their use

JOURNAL Patent: US 5714474-A 2 03-FEB-1998;

FEATURES Location/Qualifiers

source 1..3113

/organism="unknown"

BASE COUNT 961 a 684 c 586 g 882 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;

Best Local Similarity 100.0%; Pred. No. 0.27;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22

Db 1183 CGCAGTTCTGTGAACATCGACC 1162

# RESULT 5

BNCRUA/c BNCRUA/c 3113 bp DNA PLN 10-FEB-1999  
LOCUS Brassica napus crUA gene for cruciferin.  
DEFINITION X14555  
ACCESSION X14555  
VERSION X14555.1 GI:17810

KEYWORDS crUA gene; cruciferin; seed storage protein.

SOURCE rape.

ORGANISM Brassica napus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 3113)

AUTHORS Ryan, A.J.

TITLE Direct Submission

JOURNAL Submitted (02-MAR-1989) Ryan A.J., Department of Biological

Sciences, University of Durham, South Road, Durham, DH1 3LE, United

Kingdom

REFERENCE 2 (bases 1 to 3113)

AUTHORS Ryan, A.J., Royal, C.L., Hutchinson, J., and Shaw, C.H.

TITLE Genomic sequence of a 12S seed storage protein from oilseed rape

(Brassica napus c.v. jet neu)

JOURNAL Nucleic Acids Res. 17 (9), 3584 (1989)

MEDLINE 89263796

FEATURES Location/Qualifiers

source 1..3113

/organism="Brassica napus"

/strain="Jet Neuf"

/db\_xref="taxon:3708"

/tissue.type="leaf"

644..650

/note="pot. TATA-box"

680..3010

/note="primary transcript"

Join(709..991,1220..1581,2049..2870)

/codon\_start=1

/product="cruciferin"

/protein\_id="CAA32692.1"

promoter

precursor\_RNA

CDS

```
/db_xref="GI:17811"
/db_xref="SWISS-PROT:P11090"
/translation="MARLSLILSFSLALLFLHGSTAOQFPNECQLDOLNALEPSHVL
KAERGRLEVWDHAPQLRCGVSFVRVILIESKGLYLSPFFSTALSPVAKEGELMGRV
VLCATETQDSVFPQSGSPFGEQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
FRDMHQVHERHTGDTIATHPGVAQWFDNGNQPLVIVSLDLASHQNLDRNRPFF
LAGNPGQGWIEGREQQPKNLLNGFTPEVLAKAFKIDVTAQQLNQDNNGRNIIR
VQGFPSVIRPRLRSQRPOEVEVNGLEETICSARCTDNLDDPSNADVYKPOLGYISTLNS
YDLPLRLRLSALRGSIRQNAWLPQWNAANAVLVYTDGEAHVQVYVNDNGDRVFDG
QVSGQQLLSIPQGSFVVKRATSEQFRWIEFKTNANAQINTLAGRTSVLRGLPLEVIN
GYQLSLEEARVKENTITETLTHSSGPASGGPKRADA"
992..1219
/note="intron I"
intron
1582..2048
/note="intron II"
intron
2915..2920
/note="polyA signal"
misc_feature
2989..2994
/note="polyA signal"
misc_feature
BASE COUNT 961 a 685 c 586 g 881 t
ORIGIN
1 cgcagttctgtgaacatcgacc 22
|||||
Db 1183 CGCAGTTCTGTGACATCGACC 1162

RESULT 6
BNC1G/c BNC1G BNC1G 3198 bp DNA PLN 04-APR-1995
LOCUS B.napus BNC1 gene for cruciferin storage protein.
DEFINITION X59294
ACCESSION X59294
VERSION X59294.1 GI:17790
KEYWORDS cruciferin; cruciferin storage protein.
SOURCE rape.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 3198)
Breen,J.P. and Crouch,M.L.
Molecular analysis of a cruciferin storage protein gene family of
Brassica napus
Plant Mol. Biol. 19 (6), 1049-1055 (1992)
92379259
2 (bases 1 to 3198)
Breen,J.P.
Direct Submission
Submitted (24-APR-1991) Breen J.P., Jordan Hall, Indiana
University, Bloomington, IN 47401, USA
See also M16860 & X59295 (for Bnc2 gene).
FEATURES
Location/Qualifiers
1..3198
/organism="Brassica napus"
/strain="C.V. Tower"
/db_xref="taxon:3708"
/cell_line="ED8767"
/clone_lib="Ch4a"
/clone="lambda Bnc1"
join(<709..991,1220..1584,2069..2488,2653..>3057)
/gene="Bnc1"
709..3057
/gene="Bnc1"
join(709..991,1220..1584,2069..2488,2653..3057)
/gene="Bnc1"
/codon_start=1
/product="cruciferin storage protein"
/protein_id="CAA41984.1"

/db_xref="GI:762919"
/db_xref="SWISS-PROT:P33523"
/translation="MARLSLILSFSLALLFLHGSTAOQFPNECQLDOLNALEPSHVL
KAERGRLEVWDHAPQLRCGVSFVRVILIESKGLYLSPFFSTALSPVAKEGELMGRV
VPGCAETQDSVFPQSGSPFGEQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
FRDMHQVHERHTGDTIATHPGVAQWFDNGNQPLVIVSLDLASHQNLDRNRPFF
YLAGNPGQGWIEGREQQPKNLLNGFTPEVLAKAFKIDVTAQQLNQDNNGRNIIR
RVQGFPSVIRPRLRSQRPOEVEVNGLEETICSARCTDNLDDPSNADVYKPOLGYISTL
NSYDLPILRLRLSALRGSIRQNAWLPQWNAANAVLVYTDGEAHVQVYVNDNGDRVFDG
QVSGQQLLSIPQGSFVVKRATSEQFRWIEFKTNANAQINTLAGRTSVLRGLPLEVIN
SNGQISLEEARVKENTITETLTHSSGPASGGPKRADA"
<709..991
/gene="Bnc1"
exon
992..1219
/number=1
intron
/gene="Bnc1"
/number=1
exon
1220..1584
/gene="Bnc1"
intron
1585..2068
/number=2
exon
/gene="Bnc1"
/number=2
exon
2069..2488
/gene="Bnc1"
intron
2489..2652
/gene="Bnc1"
exon
2654..>3057
/number=3
exon
/gene="Bnc1"
/number=4
polyA_signal 3102..3107
polyA_signal 3176..3181
BASE COUNT 992 a 690 c 600 g 916 t
ORIGIN
Query Match 100.0%; Score 22; DB 8; Length 3198;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cgcagttctgtgaacatcgacc 22
|||||
Db 1183 CGCAGTTCTGTGACATCGACC 1162

RESULT 7
AC021599
ID AC021599 standard; DNA; HTG; 143830 BP.
XX
AC AC021599;
XX
SV AC021599.3
XX
DT 17-JAN-2000 (Rel. 62, Created)
XX
DT 21-SEP-2000 (Rel. 65, Last updated, Version 3)
XX
DE Homo sapiens clone RP11-189E18, WORKING DRAFT SEQUENCE, 29 unordered
DE pieces.
XX
KW HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP 1-143830
RA Birren B., Linton L., Nusbaum C., Lander E.;
RT "Homo sapiens, clone RP11-189E18";
RL Unpublished.
XX
```

```
/db_xref="GI:17811"
/db_xref="SWISS-PROT:P11090"
/translation="MARLSLILSFSLALLFLHGSTAOQFPNECQLDOLNALEPSHVL
KAERGRLEVWDHAPQLRCGVSFVRVILIESKGLYLSPFFSTALSPVAKEGELMGRV
VLCATETQDSVFPQSGSPFGEQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
FRDMHQVHERHTGDTIATHPGVAQWFDNGNQPLVIVSLDLASHQNLDRNRPFF
LAGNPGQGWIEGREQQPKNLLNGFTPEVLAKAFKIDVTAQQLNQDNNGRNIIR
VQGFPSVIRPRLRSQRPOEVEVNGLEETICSARCTDNLDDPSNADVYKPOLGYISTLNS
YDLPLRLRLSALRGSIRQNAWLPQWNAANAVLVYTDGEAHVQVYVNDNGDRVFDG
QVSGQQLLSIPQGSFVVKRATSEQFRWIEFKTNANAQINTLAGRTSVLRGLPLEVIN
GYQLSLEEARVKENTITETLTHSSGPASGGPKRADA"
992..1219
/note="intron I"
intron
1582..2048
/note="intron II"
intron
2915..2920
/note="polyA signal"
misc_feature
2989..2994
/note="polyA signal"
misc_feature
BASE COUNT 961 a 685 c 586 g 881 t
ORIGIN
1 cgcagttctgtgaacatcgacc 22
|||||
Db 1183 CGCAGTTCTGTGACATCGACC 1162

RESULT 6
BNC1G/c BNC1G BNC1G 3198 bp DNA PLN 04-APR-1995
LOCUS B.napus BNC1 gene for cruciferin storage protein.
DEFINITION X59294
ACCESSION X59294
VERSION X59294.1 GI:17790
KEYWORDS cruciferin; cruciferin storage protein.
SOURCE rape.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 3198)
Breen,J.P. and Crouch,M.L.
Molecular analysis of a cruciferin storage protein gene family of
Brassica napus
Plant Mol. Biol. 19 (6), 1049-1055 (1992)
92379259
2 (bases 1 to 3198)
Breen,J.P.
Direct Submission
Submitted (24-APR-1991) Breen J.P., Jordan Hall, Indiana
University, Bloomington, IN 47401, USA
See also M16860 & X59295 (for Bnc2 gene).
FEATURES
Location/Qualifiers
1..3198
/organism="Brassica napus"
/strain="C.V. Tower"
/db_xref="taxon:3708"
/cell_line="ED8767"
/clone_lib="Ch4a"
/clone="lambda Bnc1"
join(<709..991,1220..1584,2069..2488,2653..>3057)
/gene="Bnc1"
709..3057
/gene="Bnc1"
join(709..991,1220..1584,2069..2488,2653..3057)
/gene="Bnc1"
/codon_start=1
/product="cruciferin storage protein"
/protein_id="CAA41984.1"

/db_xref="GI:762919"
/db_xref="SWISS-PROT:P33523"
/translation="MARLSLILSFSLALLFLHGSTAOQFPNECQLDOLNALEPSHVL
KAERGRLEVWDHAPQLRCGVSFVRVILIESKGLYLSPFFSTALSPVAKEGELMGRV
VPGCAETQDSVFPQSGSPFGEQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
FRDMHQVHERHTGDTIATHPGVAQWFDNGNQPLVIVSLDLASHQNLDRNRPFF
YLAGNPGQGWIEGREQQPKNLLNGFTPEVLAKAFKIDVTAQQLNQDNNGRNIIR
RVQGFPSVIRPRLRSQRPOEVEVNGLEETICSARCTDNLDDPSNADVYKPOLGYISTL
NSYDLPILRLRLSALRGSIRQNAWLPQWNAANAVLVYTDGEAHVQVYVNDNGDRVFDG
QVSGQQLLSIPQGSFVVKRATSEQFRWIEFKTNANAQINTLAGRTSVLRGLPLEVIN
SNGQISLEEARVKENTITETLTHSSGPASGGPKRADA"
<709..991
/gene="Bnc1"
exon
992..1219
/number=1
intron
/gene="Bnc1"
/number=1
exon
1220..1584
/gene="Bnc1"
intron
1585..2068
/number=2
exon
/gene="Bnc1"
/number=2
exon
2069..2488
/gene="Bnc1"
intron
2489..2652
/gene="Bnc1"
exon
2654..>3057
/number=3
exon
/gene="Bnc1"
/number=4
polyA_signal 3102..3107
polyA_signal 3176..3181
BASE COUNT 992 a 690 c 600 g 916 t
ORIGIN
Query Match 100.0%; Score 22; DB 8; Length 3198;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cgcagttctgtgaacatcgacc 22
|||||
Db 1183 CGCAGTTCTGTGACATCGACC 1162

RESULT 7
AC021599
ID AC021599 standard; DNA; HTG; 143830 BP.
XX
AC AC021599;
XX
SV AC021599.3
XX
DT 17-JAN-2000 (Rel. 62, Created)
XX
DT 21-SEP-2000 (Rel. 65, Last updated, Version 3)
XX
DE Homo sapiens clone RP11-189E18, WORKING DRAFT SEQUENCE, 29 unordered
DE pieces.
XX
KW HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP 1-143830
RA Birren B., Linton L., Nusbaum C., Lander E.;
RT "Homo sapiens, clone RP11-189E18";
RL Unpublished.
XX
```

```

[2]
RN 1-143830
RA Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
RA Anderson S., Baldwin J., Barna N., Beckerly R., Bida F., Boguslavskiy L.,
RA Boukhgalter B., Brown A., Burkett G., Castle A., Choepel Y., Colangelo M.,
RA Collins S., Collymore A., Cooke P., DeArellano K., Dewar K., Domino M.,
RA Doyle M., Fenestor J., Ferreira P., FitzHugh W., Forrest C., Gage D.,
RA Galagan J., Gardyna S., Grant G., Hagos B., Heaford A., Horton L.,
RA Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,
RA Landers T., Lehoczy J., Levine R., Lieu C., Karatas A., Klein J.,
RA Macdonald P., Marquis N., McEwan P., McGuirk A., McKernan K., McPheeters R.,
RA Meldrim J., Meneus L., Morrow J., Naylor J., Norman C.H., O'Connor T.,
RA O'Donnell P., Oliver T.M., Peterson K., Pierre N., Pisani C., Pollara V.,
RA Raymond C., Riley R., Rothman D., Roy A., Santos R., Severy P., Spencer B.,
RA Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J., Testaye S.,
RA Theodore J., Tirrell A., Vassiliev H., Viel R., Vo A., Wu X., Wyman D.,
RA Ye W.J., Zimmer A., Zody M.;
RT
RL
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL Cambridge, MA 02141, USA
XX
CC On Sep 20, 2000 this sequence version replaced gi:9152655.
CC All repeats were identified using RepeatMasker:
CC Smit, A.F.A. & Green, P. (1996-1997)
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC ----- Genome Center
CC Center: Whitehead Institute/ MIT Center for Genome Research
CC Center code: W19K
CC Web site: http://www-seq.wi.mit.edu
CC Contact: sequence_submissions@genome.wi.mit.edu
CC ----- Project Information
CC Center project name: L5793
CC Center clone name: 189_E_18
CC ----- Summary Statistics
CC Sequencing vector: M13; M77815; 100% of reads
CC Chemistry: Dye-terminator Big Dye; 100% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 130936 bases at least Q40
CC Consensus quality: 137518 bases at least Q30
CC Consensus quality: 139807 bases at least Q20
CC Insert size: 150000; agarose-fp
CC Quality coverage: 3.7 in Q20 bases; agarose-fp
CC Quality coverage: 3.9 in Q20 bases; sum-of-contigs
CC -----
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 29 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC
CC 1 830: contig of 830 bp in length
CC * 831 930: gap of 100 bp
CC * 931 2470: contig of 1540 bp in length
CC * 2471 2570: gap of 100 bp
CC * 2571 4343: contig of 1773 bp in length
CC * 4344 4443: gap of 100 bp
CC * 4444 5565: contig of 1122 bp in length
CC * 5566 5665: gap of 100 bp
CC * 5666 7543: contig of 1878 bp in length
CC * 7544 7643: gap of 100 bp
CC * 7644 9898: contig of 2255 bp in length
CC * 9899 9998: gap of 100 bp
CC * 9999 11795: contig of 1797 bp in length
CC * 11796 11895: gap of 100 bp
CC * 11896 14229: contig of 2334 bp in length
CC * 14230 14329: gap of 100 bp
CC * 14330 16423: contig of 2094 bp in length
CC * 16424 16523: gap of 100 bp
CC * 16524 19064: contig of 2541 bp in length

```

```

CC * 19065 19164: gap of 100 bp
CC * 19165 21284: contig of 2120 bp in length
CC * 21285 21384: gap of 100 bp
CC * 21385 23749: contig of 2365 bp in length
CC * 23750 23849: gap of 100 bp
CC * 23850 27606: contig of 3757 bp in length
CC * 27607 27706: gap of 100 bp
CC * 27707 31136: contig of 3430 bp in length
CC * 31137 31236: gap of 100 bp
CC * 31237 35355: contig of 4119 bp in length
CC * 35356 35455: gap of 100 bp
CC * 35456 39501: contig of 4046 bp in length
CC * 39502 39601: gap of 100 bp
CC * 39602 44277: contig of 4676 bp in length
CC * 44278 44377: gap of 100 bp
CC * 44378 47808: contig of 3431 bp in length
CC * 47809 47908: gap of 100 bp
CC * 47909 52253: contig of 4345 bp in length
CC * 52254 52353: gap of 100 bp
CC * 52354 57551: contig of 5198 bp in length
CC * 57552 57651: gap of 100 bp
CC * 57652 64108: contig of 6457 bp in length
CC * 64109 64208: gap of 100 bp
CC * 64209 71334: contig of 7126 bp in length
CC * 71335 71434: gap of 100 bp
CC * 71435 79411: contig of 7977 bp in length
CC * 79412 79511: gap of 100 bp
CC * 79512 86611: contig of 7100 bp in length
CC * 86612 86711: gap of 100 bp
CC * 86712 94930: contig of 8219 bp in length
CC * 94931 95030: gap of 100 bp
CC * 95031 101723: contig of 6693 bp in length
CC * 101724 101823: gap of 100 bp
CC * 101824 116518: contig of 14695 bp in length
CC * 116519 116618: gap of 100 bp
CC * 116619 129081: contig of 12463 bp in length
CC * 129082 129181: gap of 100 bp
CC * 129182 143830: contig of 14649 bp in length.
XX
FH
FH Key Location/Qualifiers
FH Source 1. .143830
FH /db_xref="taxon:9606"
FH /organism="Homo sapiens"
FH /clone="Rp11-189E18"
FH /clone_lib="RPC1-11 Human Male BAC"
FH
FH misc_feature 1. .830
FH /note="assembly_fragment clone_end:SP6 vector_side:left"
FH
FH misc_feature 931. .2470
FH /note="assembly_fragment"
FH
FH misc_feature 2571. .4343
FH /note="assembly_fragment"
FH
FH misc_feature 4444. .5565
FH /note="assembly_fragment"
FH
FH misc_feature 5666. .7543
FH /note="assembly_fragment"
FH
FH misc_feature 7644. .9898
FH /note="assembly_fragment"
FH
FH misc_feature 9999. .11795
FH /note="assembly_fragment"
FH
FH misc_feature 11896. .14229
FH /note="assembly_fragment"
FH
FH misc_feature 14330. .16423
FH /note="assembly_fragment"
FH
FH misc_feature 16524. .19064
FH /note="assembly_fragment"
FH
FH misc_feature 19165. .21284
FH /note="assembly_fragment"
FH
FH misc_feature 21385. .23749
FH /note="assembly_fragment"
FH
FH misc_feature 23850. .27606
FH /note="assembly_fragment"
FH
FH misc_feature 27707. .31136

```

FT misc\_feature /note="assembly\_fragment"  
31237. .35355  
FT /note="assembly\_fragment"  
35456. .39501  
FT /note="assembly\_fragment"  
39602. .44277  
FT /note="assembly\_fragment"  
44378. .47808  
FT /note="assembly\_fragment"  
47909. .52253  
FT /note="assembly\_fragment"  
52354. .57551  
FT /note="assembly\_fragment"  
57652. .64108  
FT /note="assembly\_fragment"  
64209. .71334  
FT /note="assembly\_fragment"  
71435. .79411  
FT /note="assembly\_fragment"  
79512. .86611  
FT /note="assembly\_fragment"  
86712. .94930  
FT /note="assembly\_fragment"  
95031. .101723  
FT /note="assembly\_fragment"  
101824. .116518  
FT /note="assembly\_fragment"  
116619. .129081  
FT /note="assembly\_fragment"  
129182. .143830

Query Match 79.1%; Score 17.4; DB 33; Length 143830;  
Best Local Similarity 94.7%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20  
|||||  
Db 83635 GCAGTTCTGTGAACATAGA 83653

RESULT 8  
AC092613/c AC092613 151210 bp DNA HTG 19-JUL-2001  
LOCUS Homo sapiens chromosome 7 clone RP11-189E18, WORKING DRAFT  
DEFINITION SEQUENCE, 1 unordered pieces.  
ACCESSION AC092613 AC021599  
VERSION AC092613.1 GI:14916198  
KEYWORDS HTG; HTGS\_PHASE1; HTGS-DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 151210)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 151210)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUL-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Jul 19, 2001 this sequence version replaced gi:10198499.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0189E18  
Drafting center: WIBR  
----- Summary Statistics -----

Sequencing vector: M13; 26%  
Chemistry: Dye-terminator; 37%  
Chemistry: Dye-terminator; 0% of reads  
Chemistry: Dye-terminator; Big Dye; 63% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 148642 bases at least Q40  
Consensus quality: 149695 bases at least Q30  
Consensus quality: 151204 bases at least Q20  
Insert size: 143000; agarose-fp  
Quality coverage: 8.04 in Q20 bases; agarose-fp  
Quality coverage: 7.60 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 151210: contig of 151210 bp in length.

FEATURES  
Location/Qualifiers

1..151210  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/clone="RP11-189E18"

misc\_feature  
1..151210  
/note="assembly\_name:Contig5  
clone\_end:SP6  
vector\_side:right"

BASE COUNT 45173 a 28928 c 30068 g 47041 t  
ORIGIN

Query Match 79.1%; Score 17.4; DB 2; Length 151210;  
Best Local Similarity 94.7%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20  
|||||  
Db 17340 GCAGTTCTGTGAACATAGA 17322

RESULT 9

AC005047 AC005047 154016 bp DNA PRI 07-OCT-2000  
LOCUS Homo sapiens BAC clone CTB-14E15 from 7q22-q32, complete sequence.  
DEFINITION AC005047  
ACCESSION AC005047  
VERSION AC005047.3 GI:10716665  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 154016)  
AUTHORS Sulston,J.E. and Waterston,R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
REFERENCE 2 (bases 1 to 154016)  
AUTHORS Courtney,L., Gillam,B., Stoneking,T., Elliott,G. and Langston,Y.  
TITLE The sequence of Homo sapiens BAC clone CTB-14E15  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 154016)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 154016)  
AUTHORS Waterston,R.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (07-OCT-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 7, 2000 this sequence version replaced gi:7631121.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_RG014E15  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send  
<mailto:egreen@ngri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-14E15 is from the first release of the human BAC library  
CTB-978SK-B. The library contains cloned DNA from the male  
fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl.  
Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8  
(1996). This clone is available from Research Genetics, Inc.  
(<http://www.resgen.com>).  
VECTOR: pBelOBAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is GS1-34D21, 200 base pair  
overlap. Actual start of this clone is at base position 1 of  
CTB-14E15; actual end is at base position 7330 of GS1-34D21.

Due to low quality, the fidelity of the sequence from 130932 to  
130982 can not be guaranteed.

FEATURES  
source

Location/Qualifiers  
1. .154016  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q22-q32"  
/clone="CTB-14E15"  
/clone\_lib="CTB-978SK-B"

repeat\_region  
589. .874  
/rpt\_family="L1"  
repeat\_region  
903. .2491  
/rpt\_family="ERV1"  
repeat\_region  
2495. .2700  
/rpt\_family="L1"  
repeat\_region  
2720. .2838  
/rpt\_family="Alu"  
repeat\_region  
2839. .3289  
/rpt\_family="L1"  
repeat\_region  
3411. .3476  
/rpt\_family="MaLR"

repeat\_region  
3477. .3783  
/rpt\_family="Alu"  
repeat\_region  
3784. .4135  
/rpt\_family="MaLR"  
repeat\_region  
4849. .5153  
/rpt\_family="L1"  
repeat\_region  
5740. .6039  
/rpt\_family="Alu"  
repeat\_region  
6077. .6151  
/rpt\_family="MER99"  
repeat\_region  
6319. .6616  
/rpt\_family="CRI"  
repeat\_region  
8383. .8761  
/rpt\_family="Alu"  
repeat\_region  
8771. .8906  
/rpt\_family="Alu"  
repeat\_region  
9238. .9514  
/rpt\_family="Alu"  
repeat\_region  
10478. .10731  
/rpt\_family="ERV1"  
STS  
11164. .11523  
/db\_xref="GI:6120582"  
repeat\_region  
12056. .12181  
/rpt\_family="MaLR"  
repeat\_region  
12256. .12482  
/rpt\_family="MIR"  
repeat\_region  
12500. .12625  
/rpt\_family="CRI"  
repeat\_region  
12717. .12750  
/rpt\_family="CRI"  
repeat\_region  
12817. .12901  
/rpt\_family="CRI"  
repeat\_region  
13158. .13555  
/rpt\_family="ERV1"  
repeat\_region  
14352. .14508  
/rpt\_family="L1"  
repeat\_region  
14509. .14900  
/rpt\_family="MaLR"  
repeat\_region  
14901. .15114  
/rpt\_family="L1"  
repeat\_region  
15152. .15625  
/rpt\_family="ERV1"  
repeat\_region  
15626. .16096  
/rpt\_family="L1"  
repeat\_region  
16147. .16636  
/rpt\_family="ERV1"  
repeat\_region  
16640. .17690  
/rpt\_family="ERVK"  
repeat\_region  
17710. .17806  
/rpt\_family="ERV1"  
repeat\_region  
17850. .17908  
/rpt\_family="L2"  
repeat\_region  
18099. .18221  
/rpt\_family="MER103"  
STS  
18305. .18622  
/db\_xref="GI:485206"  
repeat\_region  
18833. .19179  
/rpt\_family="MER2\_type"  
repeat\_region  
19215. .19509  
/rpt\_family="Alu"  
repeat\_region  
19983. .20033  
/rpt\_family="L2"  
repeat\_region  
20109. .20370  
/rpt\_family="L2"  
repeat\_region  
20708. .20760  
/rpt\_family="MaLR"  
repeat\_region  
20933. .21089  
/rpt\_family="MaLR"  
repeat\_region  
21094. .21147  
/rpt\_family="ERV1"  
repeat\_region  
21148. .21458  
/rpt\_family="Alu"  
repeat\_region  
21459. .21613

```
repeat_region      /rpt_family="ERV1"
21626. .21980
/rpt_family="MaLR"
21985. .22223
/rpt_family="ERV1"
22241. .22547
/rpt_family="Alu"
22613. .22860
/rpt_family="MaLR"
22919. .23389
/rpt_family="ERV1"
23390. .23679
/rpt_family="Alu"
23680. .23805
/rpt_family="ERV1"
24361. .24460
/rpt_family="L2"
24491. .24642
/rpt_family="MIR"
24643. .25009
/rpt_family="MaLR"
25010. .25099
/rpt_family="MIR"
25230. .25647
/rpt_family="MaLR"
25939. .26236
/rpt_family="Alu"
28723. .29014
```

Query Match 79.1%; Score 17.4; DB 9; Length 154016;  
Best Local Similarity 94.7%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gcagttctgtgaacatcga 20  
|||||  
Db 67229 GCAGTCTGTGACATAGA 67247

```
RESULT 10
AF142615/c      AF142615      287 bp      DNA      MAM      22-JUN-2000
LOCUS
DEFINITION      Equus caballus microsatellite COR078 sequence.
ACCESSION      AF142615
VERSION      AF142615.1      GI:6318588
KEYWORDS
SOURCE      horse.
ORGANISM      Equus caballus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
TITLE      Tallmadge, R.L., Hopman, T.J., Schug, M.D., Aquadro, C.F.,
JOURNAL      Bowling, A.T., Murray, J.D., Caetano, A.R. and Antczak, D.F.
MEDLINE      Equine dinucleotide repeat loci COR061-COR080
REFERENCE      Anim. Genet. 30 (6), 462-463 (1999)
PUBMED      20078362
AUTHORS      2 (bases 1 to 287)
TITLE      Tallmadge, R.L., Hopman, T.J., Schug, M.D., Aquadro, C.F.,
JOURNAL      Bowling, A.T., Murray, J.D., Caetano, A.R. and Antczak, D.F.
MEDLINE      Direct Submission
REFERENCE      Submitted (13-APR-1999) Cornell University, J.A. Baker Institute,
PUBMED      Hungerford Hill Rd., Ithaca, NY 14853, USA
AUTHORS      Location/Qualifiers
TITLE      1..287
JOURNAL      /organism="Equus caballus"
MEDLINE      /db_xref="taxon:9796"
REFERENCE      59. .70
PUBMED      /chromosome="2"
AUTHORS      /note="microsatellite COR078"
TITLE      /rpt_type=tandem
JOURNAL      /rpt_unit-gt
FEATURES      source
1..287
Location/Qualifiers
1..287
/organism="Equus caballus"
/db_xref="taxon:9796"
/chromosome="2"
/note="microsatellite COR078"
/rpt_type=tandem
/rpt_unit-gt
```

repeat\_region 56 a 37 c 102 g 92 t

BASE COUNT 56 a 37 c 102 g 92 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 4; Length 287;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 201 CGCAGTTCTGTGACATCGTCC 180

```
RESULT 11
I56095
LOCUS
DEFINITION      I56095      587 bp      DNA      PAT      07-OCT-1997
ACCESSION      I56095
VERSION      I56095.1      GI:2476889
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 587)
AUTHORS      Parodos, K. and McCarty, J.
TITLE      Nucleic acid probes for the detection of shigella
JOURNAL      Patent: US 5648481-A 3 15-JUL-1997;
FEATURES      Location/Qualifiers
1..587
source      /organism="unknown"
BASE COUNT      138 a 149 c 139 g 161 t
ORIGIN
```

Query Match 78.2%; Score 17.2; DB 6; Length 587;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 566 CGCAGTACTGTGACCTCGATC 587

```
RESULT 12
AF190914
LOCUS
DEFINITION      AF190914      646 bp      DNA      BCT      16-JAN-2000
ACCESSION      AF190914
VERSION      AF190914.1      GI:6694242
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Plasmid Escherichia coli
REFERENCE      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
AUTHORS      Escherichia.
TITLE      1 (bases 1 to 646)
JOURNAL      Schloer, S., Riedl, S., Blass, J. and Reidl, J.
MEDLINE      Genetic rearrangements of the regions adjacent to genes encoding
REFERENCE      heat-labile enterotoxins (eltAB) of enterotoxigenic Escherichia
PUBMED      coli strains
AUTHORS      Appl. Environ. Microbiol. 66 (1), 352-358 (2000)
TITLE      2 (bases 1 to 646)
JOURNAL      Reidl, J., Schloer, S., Riedl, S. and Blass, J.
MEDLINE      Direct Submission
REFERENCE      Submitted (30-SEP-1999) Zentrum fuer Infektionsforschung,
PUBMED      University of Wuerzburg, Roentgenring 11, Wuerzburg 97070, Germany
AUTHORS      Location/Qualifiers
TITLE      1..646
JOURNAL      /organism="Escherichia coli"
MEDLINE      /plasmid="Ent"
REFERENCE      /strain="K2"
PUBMED      /db_xref="taxon:562"
AUTHORS      <1. .>646
TITLE      /gene="eltB"
JOURNAL
```

gene

BASE COUNT 182 a 134 c 149 g 175 t 6 others  
ORIGIN

/note="3' flanking sequence"

Query Match 78.2%; Score 17.2; DB 1; Length 646;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||

Db 586 CGCAGTACTGTGAACCTCGATC 607

RESULT 13  
LOCUS I56097 1174 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 5 from patent US 5648481.  
ACCESSION I56097  
VERSION I56097.1 GI:2476891  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1174)  
AUTHORS Parodos,K. and McCarty,J.  
TITLE Nucleic acid probes for the detection of shigella  
JOURNAL Patent: US 5648481-A 5 15-JUL-1997;  
FEATURES Location/Qualifiers  
source 1..1174  
/organism="unknown"

BASE COUNT 252 a 296 c 276 g 330 t 20 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 1174;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||

Db 322 CGCAGTACTGTGAACCTCGATC 343

RESULT 14  
LOCUS I56101 1188 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 9 from patent US 5648481.  
ACCESSION I56101  
VERSION I56101.1 GI:2476895  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1188)  
AUTHORS Parodos,K. and McCarty,J.  
TITLE Nucleic acid probes for the detection of shigella  
JOURNAL Patent: US 5648481-A 9 15-JUL-1997;  
FEATURES Location/Qualifiers  
source 1..1188  
/organism="unknown"

BASE COUNT 258 a 311 c 273 g 330 t 16 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 1188;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||

Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 15  
LOCUS I56099 1196 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 7 from patent US 5648481.  
ACCESSION I56099  
VERSION I56099.1 GI:2476893  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1196)  
AUTHORS Parodos,K. and McCarty,J.  
TITLE Nucleic acid probes for the detection of shigella  
JOURNAL Patent: US 5648481-A 7 15-JUL-1997;  
FEATURES Location/Qualifiers  
source 1..1196  
/organism="unknown"

BASE COUNT 259 a 298 c 272 g 353 t 14 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 1196;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||

Db 195 CGCAGTACTGTGAACCTCGATC 216

Search completed: February 25, 2002, 18:03:11  
Job time: 18599 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:42 ; Search time 716.55 Seconds  
(without alignments)  
26.322 Million cell updates/sec

Title: US-09-698-903B-14

Perfect score: 22

Sequence: 1 cgcagttctgtgaacatcgacc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	AAH25458	PCR primer for end
2	22	100.0	22	AAH25458	PCR primer B04, to
3	22	100.0	3113	AAQ07003	Cruciferin A gene.
4	17.2	78.2	587	AAQ37508	Shigella specific
5	17.2	78.2	1174	AAQ37510	E. coli repeat 1.
6	17.2	78.2	1188	AAQ37514	S. flexneri repeat
7	17.2	78.2	1196	AAQ37512	E. coli repeat 2
8	16.8	76.4	2019	AAV38672	Mus musculus SCS7
9	16.8	76.4	7316	AAH21115	C. glutamicum LP-6
10	16.4	74.5	4961	AAAC77036	Human ORFX ORF2591
11	15.8	71.8	226	AAAC12754	Human secreted pro

c 12	15.8	71.8	319608	21	AAH51601	Human chromosome 1
c 13	15.8	71.8	319608	22	AAH51601	Human schizopreni
14	15.6	70.9	545	20	AAH24719	Tobacco leaf poly
15	15.6	70.9	769	18	AAH30779	Streptococcus pneu
16	15.6	70.9	769	19	AAH98688	DNA encoding a S.
17	15.6	70.9	2108	22	AAH56484	Neisseria meningit
18	15.6	70.9	2109	22	AAH56463	Neisseria meningit
19	15.6	70.9	2111	22	AAH56458	Neisseria meningit
20	15.6	70.9	2112	21	AAH75745	DNA encoding a Nei
21	15.6	70.9	2112	21	AAH54329	Neisseria meningit
22	15.6	70.9	2112	22	AAH56459	Neisseria meningit
23	15.6	70.9	2112	22	AAH56460	Neisseria meningit
24	15.6	70.9	2112	22	AAH56461	Neisseria meningit
25	15.6	70.9	2112	22	AAH56462	Neisseria meningit
26	15.6	70.9	2112	22	AAH56466	Neisseria meningit
27	15.6	70.9	2112	22	AAH56467	Neisseria meningit
c 28	15.6	70.9	2582	7	AAH60610	Vector sequence de
c 29	15.6	70.9	2585	8	AAH70629	DNA -a sequence of
c 30	15.6	70.9	8367	21	AAH81483	N. meningitidis pa
31	15.6	70.9	21185	21	AAH63350	Streptomyces globi
32	15.6	70.9	63164	21	AAH63348	Streptomyces globi
33	15.6	70.9	349980	21	AAH21544	Neisseria meningit
34	15.6	70.9	349980	21	AAH21607	Neisseria meningit
35	15.6	70.9	1437688	21	AAH81490	N. meningitidis B
c 36	15.4	70.0	140	21	AAH06575	Human secreted pro
c 37	15.4	70.0	256	21	AAH15641	Human prostate can
c 38	15.4	70.0	429	22	AAH14743	Probe #4676 for ge
c 39	15.4	70.0	429	22	AAH36109	Probe #4795 used t
c 40	15.4	70.0	429	22	AAH04546	Probe #4537 used t
41	15.4	70.0	947	22	AAH55404	Nucleotide sequenc
42	15.4	70.0	947	22	AAH55595	Nucleotide sequenc
43	15.4	70.0	947	22	AAH57184	Apoptin-associatin
44	15.4	70.0	1131	22	AAH55405	Nucleotide sequenc
45	15.4	70.0	1131	22	AAH55596	Nucleotide sequenc

#### ALIGNMENTS

#### RESULT 1

AAH25458  
ID AAH25458 standard; DNA; 22 BP.  
XX  
AC AAH25458;  
XX  
DT 05-SEP-2001 (first entry)  
XX  
DE PCR primer for endogenous sequences in transgenic plants.  
XX  
KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
KW fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.  
XX  
OS Synthetic.  
XX  
PN WO200141558-A1.  
XX  
PD 14-JUN-2001.  
XX  
PF 06-DEC-2000; 2000WO-EP12872.  
XX  
PR 08-DEC-1999; 99US-0457037.  
XX  
PA (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
PI De Both G, De Beuckeleer M;  
XX  
DR WPI; 2001-381419/40.  
XX  
PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
PT with improved qualities, comprises a male-sterility gene and fertility  
XX restorer gene, integrated into the genome -  
XX Example 5; Page 53; 98pp; English.

XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAH5457-58 were used to amplify endogenous sequences  
 CC from transgenic plants of the invention.

XX Sequence 22 BP; 5 A; 5 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;

Best Local Similarity 100.0%; Pred. No. 0.023;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22

Db 1 cgcagttctgtgaacatcgacc 22

RESULT 2

AA07003

ID AAD07003 standard; DNA; 22 BP.

AC AAD07003;

XX 06-AUG-2001 (first entry)

XX PCR primer B04, to recognise foreign DNA and flanking sequence of MS-B2.

XX MS-B2 elite event; transgenic Brassica plant; transformation event;

XX male-sterility gene; PCR primer; ss.

XX Brassica napus.

XX WO200131042-A2.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-EF10680.

XX 29-OCT-1999; 99US-0430497.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX Weston B, De Beuckeleer M;

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -

XX Example 5; Page 33; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.

XX The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.

XX Sequence 22 BP; 5 A; 7 C; 5 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;

Best Local Similarity 100.0%; Pred. No. 0.023;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22

Db 1 cgcagttctgtgaacatcgacc 22

RESULT 3

AAQ13870/C

ID AAQ13870 standard; DNA; 3113 BP.

XX AAQ13870;

XX 09-DEC-1991 (first entry)

DE Cruciferin A gene.

XX Seed storage protein; cruA; ss.

XX Brassica napus.

XX EP49376-A.

XX 02-OCT-1991.

XX 25-MAR-1991; 91EP-0200688.

XX 25-MAR-1991; 91EP-0200688.

XX 23-MAR-1990; 90US-0498561.

XX (KONN ) GIST-BROCADES NV.

XX Pen J, Sijmons PC, Van Ooyen AJJ, Rietveld K, Verwoerd TC;

XX Quax WJ;

XX WPI; 1991-289815/40.

XX Seeds contg. enhanced enzyme levels from transgenic plants - used  
 PT for catalysing reactions, increasing nutritional values or  
 PT treating digestive disorders.

XX Example; Fig 3; 38pp; English.

XX The DNA is the genomic sequence of the seed storage protein gene  
 CC cruciferin A (cruA). It can be used in the prodn. of transgenic  
 CC plants expressing cruciferin in its seeds for use in an industrial  
 CC process. The seeds contg. the cruciferin can be used without the  
 CC need for first extracting and/or isolating the enzymes. The use of  
 CC seeds for the storage of cruciferin provides a stable vehicle which  
 CC is easily packaged and transported and easily handled during use.  
 CC See also AAQ13871-Q13877.

XX Sequence 3113 BP; 961 A; 685 C; 586 G; 881 T; 0 other;

Query Match 100.0%; Score 22; DB 12; Length 3113;

Best Local Similarity 100.0%; Pred. No. 0.049;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22

Db 1183 CGCAGTCTGTGAACATCGACC 1162

RESULT 4

AAQ37508

ID AAQ37508 standard; DNA; 587 BP.

XX AAQ37508;

XX 17-JUN-1993 (first entry)

```

XX Shigella specific fragment NT15.
DE
XX
XX Chromosome; Shigella; sonnei; probe; Enteroinvasive E. coli; EIEC;
KW virulence plasmid; detection; dysentery; ss.
XX
OS Shigella sonnei.
XX
XX WO9303187-A.
PN
XX
XX 18-FEB-1993.
PD
XX
XX 28-JUL-1992; 92WO-US06617.
PF
XX
XX 31-JUL-1991; 91US-0738800.
PR
XX
XX (STAD ) AMOCO CORP.
PA
XX
XX McCarty JM, Parodos K;
PI
XX WPI; 1993-076542/09.
DR
XX
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
PT
XX
XX Claim 3; Page 89-90; 129pp; English.
PS
XX
XX The sequences given in AAQ37506-09 fragments which were derived from
CC the chromosome of Shigella sonnei. These fragments were used to
CC design probes which are specific to Shigella and Enteroinvasive E.
CC coli (EIEC). The probes are specific to a stable region of the
CC Shigella or E. coli genome, and not the unstable virulence plasmid,
CC which means that they are more reliable than previous probes in the
CC detection of dysentery causing microbes. See also AAQ37506-35.
XX
XX Sequence 587 BP; 138 A; 149 C; 139 G; 161 T; 0 other;
SQ

```

```

PA (STAD ) AMOCO CORP.
XX
XX McCarty JM, Parodos K;
PI
XX WPI; 1993-076542/09.
DR
XX
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
PT
XX
XX Disclosure; Page 91; 129pp; English.
PS
XX
XX The sequences given in AAQ37510-14 are fragments which represent repeat
CC sequences derived from the chromosome of E. coli and S. flexneri. The
CC repeat is highly conserved and has characteristics of a transposable
CC element. Over 20 copies of the repeat sequence are found in the
CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1
CC to 3 copies in some E. coli competitors, but not in other bacterial
CC species. These fragments were used to design probes which are
CC specific to Shigella and Enteroinvasive E. coli (EIEC). The probes
CC are specific to a stable region of the Shigella or E. coli genome, and
CC not the unstable virulence plasmid, which means that they are more
CC reliable than previous probes in the detection of dysentery causing
CC microbes. See also AAQ37506-35.
XX
XX Sequence 1174 BP; 252 A; 295 C; 277 G; 330 T; 20 other;
SQ

```

```

Query Match      78.2%; Score 17.2; DB 14; Length 1174;
Best Local Similarity 86.4%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
    ||||| ||||| ||||| |||||
DB 322 cgcagtactgtgaacctcgatc 343

RESULT 6
AAQ37514
ID AAQ37514 standard; DNA; 1188 BP.
XX
XX AAQ37514;
AC
XX 17-JUN-1993 (first entry)
DT
XX
XX S. flexneri repeat 2.
DE
XX
XX Chromosome; Shigella; sonnei; flexneri; probe; Enteroinvasive; EIEC;
KW E. coli; transposable element; virulence plasmid; detection; dysentery;
KW repeat; ss.
XX
XX Shigella flexneri.
OS
XX
XX WO9303187-A.
PN
XX
XX 18-FEB-1993.
PD
XX
XX 28-JUL-1992; 92WO-US06617.
PF
XX
XX 31-JUL-1991; 91US-0738800.
PR
XX
XX (STAD ) AMOCO CORP.
PA
XX
XX McCarty JM, Parodos K;
PI
XX WPI; 1993-076542/09.
DR
XX
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
PT
XX
XX

```

PS Disclosure; Page 93; 129pp; English.

CC The sequences given in AAQ37510-14 are fragments which represent repeat  
 CC sequences derived from the chromosome of *E. coli* and *S. flexneri*. The  
 CC repeat is highly conserved and has characteristics of a transposable  
 CC element. Over 20 copies of the repeat sequence are found in the  
 CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1  
 CC to 3 copies in some *E. coli* competitors, but not in other bacterial  
 CC species. These fragments were used to design probes which are  
 CC specific to Shigella and Enteroinvasive *E. coli* (EIEC). The probes  
 CC are specific to a stable region of the Shigella or *E. coli* genome, and  
 CC not the unstable virulence plasmid, which means that they are more  
 CC reliable than previous probes in the detection of dysentery causing  
 CC microbes. See also AAQ37506-35.

XX Sequence 1188 BP; 258 A; 311 C; 273 G; 330 T; 16 other;

Query Match 78.2%; Score 17.2; DB 14; Length 1188;  
 Best Local Similarity 86.4%; Pred. No. 15;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 ||||| ||||| ||||| |||||  
 Db 195 cgcagtactgtgaacctcgatc 216

RESULT 7

AAQ37512  
 ID AAQ37512 standard; DNA; 1196 BP.

AC AAQ37512;

DT 17-JUN-1993 (first entry)

DE *E. coli* repeat 2 (2).

KW Chromosome; Shigella; sonnei; flexneri; probe; Enteroinvasive; EIEC;  
 KW *E. coli*; transposable element; virulence plasmid; detection; dysentery;  
 KW repeat; ss.

OS *Escherichia coli*.

PN WO9303187-A.

PD 18-FEB-1993.

PF 28-JUL-1992; 92WO-US06617.

PR 31-JUL-1991; 91US-0738800.

PA (STAD ) AMOCO CORP.

PI McCarty JM, Parodos K;

PT WPI; 1993-076542/09.

XX Nucleic acid probes for detection of shigella and other pathogens  
 XX - used to diagnose dysentery in non-isotopic test format and have  
 XX utility in non-isotopic test formats requiring amplification for  
 XX high sensitivity

PS Disclosure; Page 92; 129pp; English.

CC The sequences given in AAQ37510-14 are fragments which represent repeat  
 CC sequences derived from the chromosome of *E. coli* and *S. flexneri*. The  
 CC repeat is highly conserved and has characteristics of a transposable  
 CC element. Over 20 copies of the repeat sequence are found in the  
 CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1  
 CC to 3 copies in some *E. coli* competitors, but not in other bacterial  
 CC species. These fragments were used to design probes which are  
 CC specific to Shigella and Enteroinvasive *E. coli* (EIEC). The probes  
 CC are specific to a stable region of the Shigella or *E. coli* genome, and

CC not the unstable virulence plasmid, which means that they are more  
 CC reliable than previous probes in the detection of dysentery causing  
 CC microbes. See also AAQ37506-35.

XX Sequence 1196 BP; 259 A; 298 C; 272 G; 353 T; 14 other;

Query Match 78.2%; Score 17.2; DB 14; Length 1196;  
 Best Local Similarity 86.4%; Pred. No. 15;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 ||||| ||||| ||||| |||||  
 Db 195 cgcagtactgtgaacctcgatc 216

RESULT 8

AAV38672

ID AAV38672 standard; DNA; 2019 BP.

AC AAV38672;

XX 27-OCT-1998 (first entry)

DE *Mus musculus* SOCS7 gene.

KW SOCS; suppressor of cytokine signalling; PCR primer;  
 KW autoimmune disease; diagnosis; cancer; treatment;  
 KW cytokine mediated cellular responsiveness; hyperimmunity;  
 KW immunosuppression; allergies; hypertension; ss.

OS *Mus musculus*.

PH Key Location/Qualifiers  
 FT CDS 2..1054

FT /\*tag= a  
 FT /product= SOCS7 protein

PN WC9820023-A1.

PD 14-MAY-1998.

PF 31-OCT-1997; 97WO-AU00729.

PR 14-FEB-1997; 97AU-0005117.

PR 01-NOV-1996; 96AU-0003384.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Alexander WS, Hilton DJ, Metcalf D, Nicholson SE;

PI Nicola NA, Richardson RT, Starr R, Viney EM, Willson TA;

XX WPI; 1998-286854/25.

DR P-PSDB; AAW62621.

XX Suppressor of cytokine signalling proteins - useful to treat  
 PT disease, injury or abnormality involving cytokine mediated cellular  
 PT responsiveness e.g. hyperimmunity, immunosuppression, allergies and  
 PT hypertension

PS Claim 14; Page 143-144; 325pp; English.

XX The sequence is that of a gene encoding a suppressor of cytokine  
 CC signalling protein (SOCS). SOCS can be used to screen for naturally  
 CC occurring antibodies to SOCS, which may occur, e.g. in some autoimmune  
 CC diseases. Alternatively, specific antibodies can be used to  
 CC screen for SOCS, which is useful as a knowledge of SOCS levels  
 CC may be important for the diagnosis of certain cancers. Soluble  
 CC SOCS polypeptides can be used to treat disease, injury or  
 CC abnormality involving cytokine mediated cellular responsiveness,  
 CC e.g. hyperimmunity, immunosuppression, allergies and hypertension.

XX Sequence 2019 BP; 427 A; 510 C; 545 G; 528 T; 9 other;

Query Match 76.4%; Score 16.8; DB 19; Length 2019;  
 Best Local Similarity 90.0%; Pred. No. 26;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gcagttctgtgaacatcgac 21  
 |||||  
 Db 124 gcagttctgtgaccatccac 143

RESULT 9  
 AAH21115/c  
 ID AAH21115 standard; DNA; 7316 BP.  
 XX AC  
 XX AAH21115;  
 DT 05-SEP-2001 (first entry)  
 XX DE  
 XX C. glutamicum LP-6 DNA encoding tetr, teta and aada.  
 XX KW L-amino acid production; replication region; antibiotic resistance;  
 KW teta; tetracycline resistance; aada; streptomycin resistance; vitamin;  
 KW spectinomycin resistance; coryneform bacterium; D-panthothenic acid;  
 KW lysine; threonine; animal nutrition; food industry; medicine;  
 KW pharmaceutical industry; ds.  
 XX KW  
 XX OS Corynebacterium glutamicum.

FH Key Location/Qualifiers  
 FT CDS complement (1444..2013)  
 FT /\*tag= a  
 FT /product= "tetR"  
 FT 2124..3275  
 FT /\*tag= b  
 FT /product= "tetA"  
 FT 5882..6721  
 FT /\*tag= c  
 FT /product= "aada"  
 XX EPI097998-A1.  
 XX PN  
 XX 09-MAY-2001.  
 XX PD  
 XX 11-OCT-2000; 2000EP-0122056.  
 XX PF  
 XX 05-NOV-1999; 99DE-1053206.  
 XX PR  
 XX (DEGS ) DEGUSSA AG.  
 XX PA  
 XX Tauch A, Kalinowski J, Puehler A, Thierbach G;  
 XX PI  
 XX WPI: 2001-391631/42  
 XX DR  
 XX P-PSDB: AAB86252, AAB86253, AAB86254.  
 XX DR  
 XX New plasmids pTET3 and pCRY4, and their composites, useful for  
 XX PT expressing genes in coryneforms for production of amino acids, vitamins  
 XX PT and nucleotides .  
 XX PT  
 XX PS Claim 1 (1.3): Page 29-35; 46pp; German.  
 XX XX

This invention describes the novel plasmids pTET3 and pCRY4, isolated  
 CC from Corynebacterium glutamicum DSM 5816. pTET3 contains about 27.8 kb;  
 CC includes a replication region of 4539 bp (1) and a 7136 bp antibiotic  
 CC resistance region (6) containing the teta gene (resistance to  
 CC tetracycline) and the aada gene (resistance to streptomycin and  
 CC spectinomycin. pCRY4 contains about 48 kbp and includes a replication  
 CC region of 1856 bp (4). Restriction maps and sequences for (1), (4) and  
 CC (6) are reproduced. Composite plasmids derived from pTET3 and pCRY4 are  
 CC used to produce strains of coryneform bacteria that produce vitamins  
 CC (especially D-panthothenic acid); nucleotides, or L-amino acids  
 CC (particularly lysine and threonine), which are useful in animal  
 CC nutrition; the food and pharmaceutical industries, and human medicine.

CC Expression plasmids based on pTET3 or pCRY4 provide high productivity and  
 CC have unexpectedly good compatibility with known plasmids. This sequence  
 CC represents a DNA sequence containing a fragment which encodes the  
 CC Corynebacterium glutamicum LP-6 tetr, teta and aada proteins described in  
 CC the method of the invention.  
 XX  
 SQ Sequence 7316 BP; 1485 A; 2116 C; 2164 G; 1551 T; 0 other;

Query Match 76.4%; Score 16.8; DB 22; Length 7316;  
 Best Local Similarity 90.0%; Pred. No. 31;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcga 20  
 |||||  
 Db 3778 CGCAGTTCTGCGATCATCGA 3759

RESULT 10  
 AAC77036/c  
 ID AAC77036 standard; cDNA; 4961 BP.  
 XX AC  
 XX AAC77036;  
 XX AC  
 DT 08-FEB-2001 (first entry)  
 XX DT  
 XX DE Human ORF2591 polynucleotide sequence SEQ ID NO:5181.  
 XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX KW  
 OS Homo sapiens.  
 XX XX  
 XX WO2000058473-A2.  
 XX PN  
 XX 05-OCT-2000.  
 XX PD  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX PF  
 XX 31-MAR-1999; 99US-0127607.  
 XX PR  
 XX 02-APR-1999; 99US-0127636.  
 XX PR  
 XX 05-APR-1999; 99US-0127728.  
 XX PR  
 XX 30-MAR-2000; 2000US-0540763.  
 XX XX  
 XX (CURA-) CURAGEN CORP.  
 XX XX  
 XX Shimkets RA, Leach M;  
 XX PI  
 XX WPI: 2000-602362/57.  
 XX DR  
 XX P-PSDB: AAB42827.  
 XX DR  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 XX PT useful for treating e.g. cancers, proliferative disorders,  
 XX PT neurodegenerative disorders and cardiovascular disease -  
 XX XX  
 XX PS Claim 5; Page 4360-4363; 5507pp; English.  
 XX XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiatic; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy.  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 4961 BP; 1411 A; 1065 C; 1123 G; 1362 T; 0 other;

Query Match 74.5%; Score 16.4; DB 21; Length 4961;  
 Best Local Similarity 94.48; Pred. No. 48;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21  
 |||||  
 Db 2639 AGTCTGTGAACATTGAC 2622

RESULT 11  
 AAC12754/c  
 ID AAC12754 standard; CDNA; 226 BP.

XX AAC12754;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 16829.

XX Human; 5' EST; expressed sequence tag; secreted protein; CDNA isolation;  
 gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Mline Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 16829; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

XX Sequence 226 BP; 75 A; 29 C; 51 G; 67 T; 4 other;

Query Match 71.8%; Score 15.8; DB 21; Length 226;  
 Best Local Similarity 89.5%; Pred. No. 62;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 cagttctgtgaacatcgac 21  
 |||||  
 Db 138 CAGTCTGTGAATTTCGAC 120

RESULT 12  
 AAH51601/c  
 ID AAH51601 standard; DNA; 319608 BP.

XX AC AAH51601;

XX 29-AUG-2001 (first entry)

XX Human chromosome 13q31-q33 genomic nucleotide sequence.

XX sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;  
 KW biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.

XX Homo sapiens.

XX WO200058510-A2.

XX 05-OCT-2000.

XX 30-MAR-2000; 2000WO-IB00435.

XX 30-MAR-1999; 99US-0126903.

XX 30-APR-1999; 99US-0131971.

XX 30-APR-1999; 99US-0132065.

XX 14-JUL-1999; 99US-0143928.

XX 27-JUL-1999; 99US-0145915.

XX 29-JUL-1999; 99US-0146452.

XX 29-JUL-1999; 99US-0146453.

XX 28-OCT-1999; 99US-0162288.

XX (GEST ) GENSET.

XX Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihaïn B;

XX Essioux L;

XX WPI; 2000-619082/59.

XX Polynucleotides comprising sequences from sbg1 and g35018 biallelic  
 PT markers are used for genotyping and detecting schizophrenia or bipolar  
 PT disorder and predisposition to these disorders -

XX Claim 1; Page 409-493; 737pp; English.

XX AAH51601 represents a human genomic nucleotide sequence comprising sbg1,  
 CC g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the  
 CC human chromosome 13q31-q33 locus. The nucleotide sequences contain  
 CC biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and  
 CC AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein  
 CC products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018  
 CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used  
 CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are  
 CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of  
 CC amplicons which comprise biallelic markers located on the chromosome  
 CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Biallelic markers



are represented in the sequences by degenerate/undefined base codes. PCR primers AAH51818 and AAH51819 are used in the isolation of sequences of the invention. The biallelic marker containing nucleotide sequences are used to determine the identity of the nucleotide at a biallelic marker in a sample DNA sequence. The nucleotide sequences may be labelled and used for genotyping by determining the identity of a nucleotide at a Region D-related biallelic marker in a biological sample from single or multiple subjects. By determining the frequency of a biallelic marker in a population an association between a genotype and a trait, a haplotype and a trait and a phenotype and a trait can be detected. The sequences can be used to determine a predisposition to or early onset of schizophrenia or bipolar disorder or a beneficial response to or side effects related to treatment against schizophrenia or bipolar disorder.

Sequence 319608 BP; 101600 A; 56677 C; 58335 G; 102722 T; 274 other;  
Query Match 71.88; Score 15.8; DB 21; Length 319608;  
Best Local Similarity 85.58; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gcagttctgtgaacatcga 20  
||||||| ||||||| |  
Db 164729 GCAGTTCCTGACATCTA 164711

RESULT 13  
AAS09301/c  
ID AAS09301 standard; DNA; 319608 BP.  
XX AAS09301;  
AC AAS09301;  
XX 26-SEP-2001 (first entry)  
XX Human schizophrenia associated gene g35030 and biallelic markers Al-A71.  
DE Human; g35030; biallelic marker; Al-A71; chromosome 13q31-q33;  
KW schizophrenia; bipolar disorder; ds.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH primer\_bind 7938..7958  
FT /\*tag= a  
FT /note= "Binds primer 99-27943.rp"  
FT 8297..8315  
FT /\*tag= b  
FT /note= "Binds primer 99-27943-150.mis"  
FT 8304..8328  
FT /\*tag= c  
FT /bound\_moiety= Probe\_99-27943-150  
FT 8316  
FT /\*tag= d  
FT /note= "Biallelic marker A1"  
FT complement (8317..8335)  
FT /\*tag= e  
FT /note= "Binds primer 99-27943-150.mis complement"  
FT complement (8446..8465)  
FT /\*tag= f  
FT /note= "Binds primer 99-27943.pu complement"  
FT 21365..21385  
FT /\*tag= g  
FT /note= "Binds primer 99-27935.rp"  
FT 21653..21671  
FT /\*tag= h  
FT /note= "Binds primer 99-27935-193.mis"  
FT 21660..21684  
FT /\*tag= i  
FT /bound\_moiety= Probe\_99-27935-193  
FT 21672  
FT /\*tag= j  
FT /note= "Biallelic marker A2"  
FT complement (21673..21691)

FT primer\_bind /\*tag= k  
FT /note= "Binds primer 99-27935-193.mis complement"  
FT complement (21845..21864)  
FT /\*tag= l  
FT /note= "Binds primer 99-27935.pu complement"  
FT 65463..65471  
FT /\*tag= m  
FT /note= "Binds primer 8-128.pu"  
FT 65466..65484  
FT /\*tag= n  
FT /note= "Binds primer 8-128-33.mis"  
FT 65473..65497  
FT /\*tag= o  
FT /bound\_moiety= Probe\_8-128-33  
FT 65485  
FT /\*tag= p  
FT /note= "Biallelic marker A3"  
FT complement (65486..65504)  
FT /\*tag= q  
FT /note= "Binds primer 8-128-33.mis complement"  
FT complement (65856..65874)  
FT /\*tag= r  
FT /note= "Binds primer 8-128.rp complement"  
FT 95034..95053  
FT /\*tag= s  
FT /note= "Binds primer 99-31960.pu"  
FT 95377..95395  
FT /\*tag= t  
FT /note= "Binds primer 99-31960-363.mis"  
FT 95384..95408  
FT /\*tag= u  
FT /bound\_moiety= Probe\_99-31960-363  
FT 95396  
FT /\*tag= v  
FT /note= "Biallelic marker A4"  
FT complement (95397..95415)  
FT /\*tag= w  
FT /note= "Binds primer 99-31960-363.mis complement"  
FT complement (95543..95563)  
FT /\*tag= x  
FT /note= "Binds primer 99-31960.rp complement"  
FT 107022..107040  
FT /\*tag= y  
FT /note= "Binds primer 99-24656.pu"  
FT 107262..107280  
FT /\*tag= z  
FT /note= "Binds primer 99-24656-260.mis"  
FT 107269..107293  
FT /\*tag= aa  
FT /bound\_moiety= Probe\_99-24656-260  
FT 107281  
FT /\*tag= ab  
FT /note= "Biallelic marker A5"  
FT complement (107282..107300)  
FT /\*tag= ac  
FT /note= "Binds primer 99-24656-260.mis complement"  
FT complement (107495..107513)  
FT /\*tag= ad  
FT /note= "Binds primer 99-24656.rp complement"  
FT 160279..160298  
FT /\*tag= ae  
FT /note= "Binds primer 99-24639.rp"  
FT 160621..160639  
FT /\*tag= af  
FT /note= "Binds primer 99-24639-163.mis"  
FT 160628..160652  
FT /\*tag= ag  
FT /bound\_moiety= Probe\_99-24639-163  
FT 160640  
FT /\*tag= ah  
FT /note= "Biallelic marker A6"  
FT complement (160641..160659)  
FT /\*tag= ai

```

FT primer_bind /note= "Binds primer 99-24639-163.mis complement"
FT 160770..160787
FT /tag= aj
FT /note= "Binds primer 99-24634.pu"
FT complement (160785..160802)
FT /tag= ak
FT /note= "Binds primer 99-24639.pu complement"
FT 160857..160875
FT /tag= al
FT /note= "Binds primer 99-24634-108.mis"
FT 160864..160888
FT /tag= am
FT /bound_moiety= Probe_99-24634-108
FT 160876
FT /tag= an
FT /note= "Biallelic marker A7"
FT complement (160877..160895)
FT /tag= ao
FT /note= "Binds primer 99-24634-108.mis complement"
FT complement (161240..161257)
FT /tag= ap
FT /note= "Binds primer 99-24634.rp complement"
FT 168813..168830
FT /tag= aq
FT /note= "Binds primer 99-7652.pu"
FT 168955..168973
FT /tag= ar
FT /note= "Binds primer 99-7652-162.mis"
FT 168962..168986
FT /tag= as
FT /bound_moiety= Probe_99-7652-162
FT 168974
FT /tag= at
FT /note= "Biallelic marker A8"
FT complement (168975..168993)
FT /tag= au
FT /note= "Binds primer 99-7652-162.mis complement"
FT complement (169331..169351)
FT /tag= av
FT /note= "Binds primer 99-7652.rp complement"
FT 170666..170686
FT /tag= aw
FT /note= "Binds primer 99-16100.pu"
FT 170791..170809
FT /tag= ax
FT /note= "Binds primer 99-16100-147.mis"
FT 170798..170822
FT /tag= ay
FT /bound_moiety= Probe_99-16100-147
FT 170810
FT /tag= az
FT /note= "Biallelic marker A9"
FT complement (170811..170829)
FT /tag= ba
FT /note= "Binds primer 99-16100-147.mis complement"
FT complement (171153..171173)
FT /tag= bb
FT /note= "Binds primer 99-16100.rp complement"
FT 173085..173085
FT /tag= bc
FT /note= "Binds primer 99-5862.rp"
FT 173339..173357
FT /tag= bd
FT /note= "Binds primer 99-5862-167.mis"
FT 173346..173370
FT /tag= be
FT /bound_moiety= Probe_99-5862-167
FT 173358
FT /tag= bf
FT /note= "Biallelic marker A10"
FT complement (173359..173377)
FT /tag= bg
FT /note= "Binds primer 99-5862-167.mis complement"

```

```

FT primer_bind complement (173495..173514)
FT /tag= bh
FT /note= "Binds primer 99-5862.pu complement"
FT 189753..189771
FT /tag= bi
FT /note= "Binds primer 99-5919.pu"
FT 189938..189956
FT /tag= bj
FT /note= "Binds primer 99-5919-215.mis"

Query Match 71.8%; Score 15.8; DB 22; Length 319608;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20
Db 164729 GCAGTTCTCTGAACATCTA 164711

RESULT 14
AAx24719
ID AAx24719 standard; cDNA; 545 BP.
XX
AC AAx24719;
XX
DT 21-JUN-1999 (first entry)
XX
DE Tobacco leaf polyphenol oxidase cDNA clone TOBPP06.
XX
KW Polyphenol oxidase; banana; tobacco; pineapple; transgenic plant;
KW vaccine; browning; spoilage; ds.
XX
OS Nicotiana tabacum.
PN WO9853080-A1.
XX
PD 26-NOV-1998.
XX
PF 19-MAY-1998; 98WO-AU00362.
PR 19-MAY-1997; 97AU-0006849.
XX
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI Robinson SP;
XX
DR WPI: 1999-070152/06.
DR P-PSDB; AAW97990.
XX
PT Nucleic acid encoding polyphenol oxidase from banana, tobacco and
PT pineapple - useful for, e.g. increasing levels of the enzyme
PT expression, which is responsible for browning and spoilage of fruits
PT after injury or damage
XX
PS Example 2; Fig 5; 47pp; English.
XX
CC This partial cDNA clone, termed TOBPP06, encodes a polyphenol oxidase
CC (PPO) polypeptide of tobacco (see also AAW97990). The clone was
CC obtained by PCR amplification of young tobacco leaf cDNA using
CC primers (see AAx24708-12) based on conserved copper binding sites of
CC plant PPOs, and identified on the basis of homology to known plant
CC PPO genes. PPO is the major enzyme responsible for browning and
CC spoilage of fruits and vegetables after damage or injury. Sense
CC PPO nucleic acids can be used to increase the levels of PPO in a
CC plant. Optionally modified sense sequences and antisense sequences
CC can be used to reduce PPO levels (by co-suppression in the case of
CC sense sequences). Transgenic plants including antisense PPO
CC sequences, and a plant vaccine including a nucleic acid encoding
CC banana, tobacco or pineapple PPO or an antisense sequence are
CC claimed.
XX
SQ Sequence 545 BP; 171 A; 115 C; 115 G; 144 T; 0 other;

```

Query Match 70.9%; Score 15.6; DB 20; Length 545;  
Best Local Similarity 81.8%; Pred. No. 90;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
    |||||  
Db 124 cgcggtctgtgaatcatggacc 145

RESULT 15

AAX30779  
ID AAX30779 standard; DNA; 769 BP.  
XX  
AC AAX30779;  
XX  
DT 20-MAY-1999 (first entry)  
XX  
DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:56.  
XX  
KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
KW streptococcal infection; pneumococcal; ss.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9737026-A1.  
XX  
PD 09-OCT-1997.  
XX  
PF 01-APR-1997; 97WO-US05306.  
XX  
PR 22-AUG-1996; 96US-0025788.  
PR 02-APR-1996; 96US-0014690.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
PI Stodola RK;  
XX  
WPI: 1997-503111/46.  
DR P-PSDB; AAY11181.  
XX  
Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
vaccines, drug screening, etc  
PT  
XX  
Claim 5; Page 91; 354pp; English.  
XX  
AAX30724 to AAX30946 represent genomic DNA sequences isolated from  
Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
encode the novel proteins given in AAY11114 to AAY11367. The proteins,  
isolated from Streptococcus pneumoniae, can be used in vaccines against  
streptococcal infections and in assays for identifying compounds that  
inhibit or activate the activity of the proteins. The antagonists can  
be used to treat an individual having need to inhibit a bacterial  
protein. Vectors expressing the proteins can be used to induce a  
protective immune response in mammals.  
XX  
SQ Sequence 769 BP; 204 A; 202 C; 140 G; 223 T; 0 other;

Query Match 70.9%; Score 15.6; DB 18; Length 769;  
Best Local Similarity 81.8%; Pred. No. 95;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
    |||||  
Db 5 cgcagttctgttaccacaaagacc 26

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:55 ; Search time 301.6 seconds  
(without alignments)  
16.520 Million cell updates/sec

Title: US-09-698-903B-14  
Perfect score: 22  
Sequence: 1 cgcagttctgtgaacatcgacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	100.0	3113	1 US-08-146-422-20	Sequence 20, Appl
C 2	22	100.0	3113	1 US-08-628-534-2	Sequence 2, Appl
C 3	17.2	78.2	587	1 US-08-375-241-3	Sequence 3, Appl
C 4	17.2	78.2	587	5 PCT-US92-06617A-3	Sequence 3, Appl
C 5	17.2	78.2	1174	1 US-08-375-241-5	Sequence 5, Appl
C 6	17.2	78.2	1174	5 PCT-US92-06617A-5	Sequence 5, Appl
C 7	17.2	78.2	1188	1 US-08-375-241-9	Sequence 9, Appl
C 8	17.2	78.2	1188	5 PCT-US92-06617A-9	Sequence 9, Appl
C 9	17.2	78.2	1196	1 US-08-375-241-7	Sequence 7, Appl
C 10	17.2	78.2	1196	5 PCT-US92-06617A-7	Sequence 7, Appl
C 11	15.2	69.1	195	1 US-08-158-189-9	Sequence 9, Appl
C 12	15.2	69.1	1126	1 US-08-233-788A-48	Sequence 48, Appl
C 13	15.2	69.1	1576	3 US-08-689-974-2	Sequence 2, Appl
C 14	15.2	69.1	1576	3 US-09-058-376-2	Sequence 2, Appl
C 15	15.2	69.1	2048	1 US-07-602-608-11	Sequence 11, Appl
C 16	15.2	69.1	2048	1 US-08-261-578-11	Sequence 11, Appl
C 17	15.2	69.1	2158	1 US-07-602-608-1	Sequence 1, Appl
C 18	15.2	69.1	2158	1 US-08-261-578-1	Sequence 1, Appl
C 19	14.8	67.3	30	1 US-07-602-608-18	Sequence 18, Appl
C 20	14.8	67.3	30	1 US-08-261-578-18	Sequence 18, Appl
C 21	14.8	67.3	425	4 US-09-328-111-19	Sequence 19, Appl
C 22	14.6	66.4	952	4 US-09-174-768-2	Sequence 2, Appl
C 23	14.6	66.4	1062	2 US-08-468-819-92	Sequence 92, Appl
C 24	14.6	66.4	3330	1 US-08-149-103-1	Sequence 1, Appl
C 25	14.6	66.4	3330	1 US-08-451-883-1	Sequence 1, Appl
C 26	14.6	66.4	3656	1 US-08-393-734-1	Sequence 1, Appl
C 27	14.6	66.4	3656	4 US-08-894-489-1	Sequence 1, Appl

28	14.6	66.4	4649	6	5183745-1	Patent No. 5183745
29	14.6	66.4	6443	6	5183745-5	Patent No. 5183745
C 30	14.6	66.4	9592	1	US-08-393-734-3	Sequence 3, Appl
C 31	14.6	66.4	9592	4	US-08-894-489-3	Sequence 3, Appl
C 32	14.6	66.4	80161	3	US-09-036-987A-1	Sequence 1, Appl
C 33	14.6	66.4	80161	4	US-09-370-700-1	Sequence 1, Appl
34	14.4	65.5	1929	5	PCT-US93-00031-18	Sequence 18, Appl
35	14.4	65.5	1932	5	PCT-US93-00031-20	Sequence 20, Appl
36	14.4	65.5	1941	5	PCT-US93-00031-10	Sequence 10, Appl
37	14.4	65.5	1941	5	PCT-US93-00031-22	Sequence 22, Appl
38	14.4	65.5	2205	5	PCT-US93-00031-12	Sequence 12, Appl
39	14.4	65.5	2208	5	PCT-US93-00031-14	Sequence 14, Appl
40	14.4	65.5	2217	5	PCT-US93-00031-8	Sequence 8, Appl
41	14.4	65.5	2220	5	PCT-US93-00031-16	Sequence 16, Appl
42	14.4	65.5	2811	4	US-08-482-073-3	Sequence 3, Appl
43	14.4	65.5	2813	2	US-08-344-155C-99	Sequence 99, Appl
44	14.4	65.5	2813	4	US-09-009-490A-90	Sequence 90, Appl
45	14.4	65.5	3080	4	US-08-482-073-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-08-146-422-20/c  
; Sequence 20, Application US/08146422  
; Patent No. 5543576  
; GENERAL INFORMATION:  
; APPLICANT: VAN OOLJEN, ALBERT J. J.  
; APPLICANT: RIETVELD, KRIJN  
; APPLICANT: HOEKEMA, ANDREAS  
; APPLICANT: PEN, JAN  
; APPLICANT: SIJMONS, PETER C.  
; APPLICANT: VERWOERD, TEUNIS C.  
; APPLICANT: QUAX, WILHEMUS J.  
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/146,422  
; FILING DATE: 02-NOV-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KENNEDY, BILL  
; REGISTRATION NUMBER: 33,407  
; REFERENCE/DOCKET NUMBER: 44615-20011.23  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3113 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-146-422-20

Query Match 100.0%; Score 22; DB 1; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.012;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 1183 CGCAGTTCTGTGAACATCGACC 1162

## RESULT 2

US-08-626-554-2/c  
; Sequence 2, Application US/08626554  
; Patent No. 5714474  
; GENERAL INFORMATION:  
; APPLICANT: VAN OIJEN, ALBERT J.J.  
; APPLICANT: RIETVELD, KRIJN  
; APPLICANT: HOEKEMA, ANDREAS  
; APPLICANT: PEN, JAN  
; APPLICANT: SIJMONS, PETER C.  
; APPLICANT: VERWOERD, TEUNIS C.  
; APPLICANT: QUAK, WILHEMUS J.  
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
; TITLE OF INVENTION: USE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 PENNSYLVANIA AVENUE NW  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/626,554  
; FILING DATE: 02-APR-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 26192-20011.10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030 MRSNFOERSWSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3113 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-626-554-2

Query Match 100.0%; Score 22; DB 1; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.012;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 1183 CGCAGTTCTGTGAACATCGACC 1162

## RESULT 3

US-08-375-241-3  
; Sequence 3, Application US/08375241  
; Patent No. 5648481  
; GENERAL INFORMATION:  
; APPLICANT: Parodos, Kyriaki  
; APPLICANT: McCarty, Janice  
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
; TITLE OF INVENTION: Shigella

; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/375,241  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/738,800  
; FILING DATE: 31-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: GTR90-04  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 587 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-375-241-3

Query Match 78.2%; Score 17.2; DB 1; Length 587;  
Best Local Similarity 86.4%; Pred. No. 3.1;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 566 CGCAGTACTGTGAACCTCGATC 587

## RESULT 4

PCT-US92-06617A-3  
; Sequence 3, Application PC/TUS9206617A  
; GENERAL INFORMATION:  
; APPLICANT: Parodos, Kyriaki  
; APPLICANT: McCarty, Janice  
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
; TITLE OF INVENTION: Shigella  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amoco Corporation  
; STREET: 200 East Randolph Drive, P.O. Box 87703  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60680  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06617A  
; FILING DATE: 19920728  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER: US 07/738,800  
; FILING DATE: 31-JUL-1991  
; ATTORNEY/AGENT INFORMATION:

NAME: Galloway, Norval B.  
REGISTRATION NUMBER: 33,595  
REFERENCE/DOCKET NUMBER: GTR90-04 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-856-7180  
TELEFAX: 312-856-4972  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 587 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
PCT-US92-06617A-3

Query Match 78.2%; Score 17.2; DB 5; Length 587;  
Best Local Similarity 86.4%; Pred. No. 3.1;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 566 CGCAGTACTGTGAACCTCGATC 587

RESULT 5  
US-08-375-241-5  
Sequence 5, Application US/08375241  
Patent No. 5648481  
GENERAL INFORMATION:  
APPLICANT: Parodos, Kyriaki  
APPLICANT: McCarty, Janice  
TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
TITLE OF INVENTION: Shigella  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02173

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/375,241  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/738,800  
FILING DATE: 31-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: GTR90-04  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1174 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-375-241-5

Query Match 78.2%; Score 17.2; DB 1; Length 1174;  
Best Local Similarity 86.4%; Pred. No. 3.6;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 322 CGCAGTACTGTGAACCTCGATC 343

RESULT 6  
PCT-US92-06617A-5  
Sequence 5, Application PC/TUS9206617A  
GENERAL INFORMATION:  
APPLICANT: Parodos, Kyriaki  
APPLICANT: McCarty, Janice  
TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
TITLE OF INVENTION: Shigella  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corporation  
STREET: 200 East Randolph Drive, P.O. Box 87703  
CITY: Chicago  
STATE: Illinois  
COUNTRY: U.S.A.  
ZIP: 60680  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/06617A  
FILING DATE: 19920728  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/738,800  
FILING DATE: 31-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, Norval B.  
REGISTRATION NUMBER: 33,595  
REFERENCE/DOCKET NUMBER: GTR90-04 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-856-7180  
TELEFAX: 312-856-4972  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1174 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
PCT-US92-06617A-5

Query Match 78.2%; Score 17.2; DB 5; Length 1174;  
Best Local Similarity 86.4%; Pred. No. 3.6;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 322 CGCAGTACTGTGAACCTCGATC 343

RESULT 7  
US-08-375-241-9  
Sequence 9, Application US/08375241  
Patent No. 5648481  
GENERAL INFORMATION:  
APPLICANT: Parodos, Kyriaki  
APPLICANT: McCarty, Janice  
TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
TITLE OF INVENTION: Shigella  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: U.S.A.

;;  
;; ZIP: 02173  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/375,241  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/738,800  
;; FILING DATE: 31-JUL-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Granahan, Patricia  
;; REGISTRATION NUMBER: 32,227  
;; REFERENCE/DOCKET NUMBER: GTR90-04  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-861-6240  
;; TELEFAX: 617-861-9540  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1188 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;;  
;; US-08-375-241-9

Query Match 78.2%; Score 17.2; DB 1; Length 1188;  
Best Local Similarity 86.4%; Pred. No. 3.6;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 8  
PCT-US92-06617A-9  
;; Sequence 9, Application PC/TUS9206617A  
;; GENERAL INFORMATION:  
;; APPLICANT: Parodos, Kyriaki  
;; APPLICANT: McCarthy, Janice  
;; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
;; TITLE OF INVENTION: Shigella  
;; NUMBER OF SEQUENCES: 30  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Amoco Corporation  
;; STREET: 200 East Randolph Drive, P.O. Box 87703  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: U.S.A.  
;; ZIP: 60680  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US92/06617A  
;; FILING DATE: 19920728  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/738,800  
;; FILING DATE: 31-JUL-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Galloway, Norval B.  
;; REGISTRATION NUMBER: 33,595  
;; REFERENCE/DOCKET NUMBER: GTR90-04 PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312-856-7180  
;; TELEFAX: 312-856-4972  
;; INFORMATION FOR SEQ ID NO: 9:

;;  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1188 base pairs  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;;  
;; PCT-US92-06617A-9

Query Match 78.2%; Score 17.2; DB 5; Length 1188;  
Best Local Similarity 86.4%; Pred. No. 3.6;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 9  
US-08-375-241-7  
;; Sequence 7, Application US/08375241  
;; Patent No. 5648481  
;; GENERAL INFORMATION:  
;; APPLICANT: Parodos, Kyriaki  
;; APPLICANT: McCarthy, Janice  
;; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
;; TITLE OF INVENTION: Shigella  
;; NUMBER OF SEQUENCES: 30  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
;; STREET: Two Militia Drive  
;; CITY: Lexington  
;; STATE: Massachusetts  
;; COUNTRY: U.S.A.  
;; ZIP: 02173  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/375,241  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/738,800  
;; FILING DATE: 31-JUL-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Granahan, Patricia  
;; REGISTRATION NUMBER: 32,227  
;; REFERENCE/DOCKET NUMBER: GTR90-04  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-861-6240  
;; TELEFAX: 617-861-9540  
;; INFORMATION FOR SEQ ID NO: 7:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1196 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;;  
;; US-08-375-241-7

Query Match 78.2%; Score 17.2; DB 1; Length 1196;  
Best Local Similarity 86.4%; Pred. No. 3.6;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||  
Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 10  
PCT-US92-06617A-7



```

; Sequence 7, Application PC/TUS9206617A
; GENERAL INFORMATION:
; APPLICANT: Patodos, Kyriaki
; APPLICANT: McCarthy, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corporation
; STREET: 200 East Randolph Drive, P.O. Box 87703
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06617A
; FILING DATE: 19920728
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, Norval B.
; REGISTRATION NUMBER: 33,595
; REFERENCE/DOCKET NUMBER: GTR90-04 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-856-7180
; TELEFAX: 312-856-4972
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US92-06617A-7

```

```

Query Match 78.2%; Score 17.2; DB 5; Length 1196;
Best Local Similarity 86.4%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatgcacc 22
Db 195 CGCAGTACTGTGACCTCGATC 216

RESULT 11
US-08-158-189-9
; Sequence 9, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides.
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/158,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,232
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnson, Phillip S.
; REGISTRATION NUMBER: 27,200
; REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-158-189-9

Query Match 69.1%; Score 15.2; DB 1; Length 195;
Best Local Similarity 85.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcga 20
Db 130 CACAGTTCAGTGAGCATCGA 149

RESULT 12
US-08-233-788A-48/c
; Sequence 48, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1126 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

```
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(3..23, 27..944, 948..1124)
US-08-233-788A-48

Query Match          69.1%; Score 15.2; DB 1; Length 1126;
Best Local Similarity 85.0%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcgacc 21
   ||||| ||||| ||||| |||||
Db 492 gctgttcggtgaacatcgcc 473

RESULT 13
US-08-689-974-2/c
; Sequence 2, Application US/08689974
; Patent No. 5776732
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Murray, Lynn E.
; TITLE OF INVENTION: NOVEL HUMAN INDUCED TUMOR PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,974
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0113 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT03
; CLONE: 530522
; US-08-689-974-2

Query Match          69.1%; Score 15.2; DB 1; Length 1576;
Best Local Similarity 85.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cagttctgtgaacatcgacc 22
   ||||| ||||| ||||| |||||
Db 893 CAGCTCGGTGAACATCGTCC 874

RESULT 14
US-09-698-974-2/c
; Sequence 2, Application US/09058376
; Patent No. 5382524
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; ACTIVE ALPHA-N-ACETYL GALACTOSAMINIDASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; US-09-698-974-2
```

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/602,608  
; FILING DATE: 24-OCT-1990  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 6923-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2048 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; US-07-602-608-11

Query Match 69.1%; Score 15.2; DB 1; Length 2048;  
Best Local Similarity 85.0%; Pred. No. 46;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcga 20  
||||| ||||| |||||  
Db 101 CGCAGGTCGTGGACATCAA 82

Search completed: February 25, 2002, 18:05:56  
Job time: 18604 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:21 : Search time 8261.74 Seconds  
(without alignments)  
28.615 Million cell updates/sec

Title: US-09-698-903B-14

Perfect score: 22  
Sequence: 1 cgcagttctgtgaacatgacc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	80.9	653	10	BE377681 601229824
2	16.8	76.4	244	10	AI561687 vw88f09.x
3	16.8	76.4	279	13	AZ777284 2M0011K06
4	16.8	76.4	346	10	AI019782 ua90h12.r
5	16.8	76.4	605	13	AL026432 Fugu rubr
6	16.8	76.4	611	13	AL026427 Fugu rubr
7	16.8	76.4	613	13	AZ777336 2M0011I08
8	16.8	76.4	1301	12	AK006515 Mus muscu
9	16.4	74.5	206	10	AW086177 xc77ell.x
10	16.4	74.5	234	10	AA701011 zg55d04.s
11	16.4	74.5	270	11	F02043 HSCOMG102 n
12	16.4	74.5	290	10	AA569516 nf23e01.s

13	16.4	74.5	303	11	F03921
14	16.4	74.5	304	10	AA843570
15	16.4	74.5	335	10	AI084952
16	16.4	74.5	399	10	AI658711
17	16.4	74.5	412	11	BF066458
18	16.4	74.5	435	11	N49244
19	16.4	74.5	441	10	AW196345
20	16.4	74.5	455	11	BF483372
21	16.4	74.5	460	10	AA679411
22	16.4	74.5	463	10	AA426087
23	16.4	74.5	483	10	AI149964
24	16.4	74.5	487	10	AI802194
25	16.4	74.5	499	10	AA421463
26	16.4	74.5	508	10	AA700220
27	16.4	74.5	526	10	AA235369
28	16.4	74.5	545	11	R60174
29	16.4	74.5	553	10	AI631483
30	16.4	74.5	560	10	AW013257
31	16.4	74.5	582	11	W91993
32	16.4	74.5	612	13	BH121721
33	16.4	74.5	630	10	AA700001
34	16.4	74.5	650	10	AA952652
35	16.4	74.5	673	10	AA868429
36	16.4	74.5	679	10	AI680857
37	16.4	74.5	706	13	AZ717353
38	16.4	74.5	758	10	AI740628
39	16.4	74.5	768	10	AI637690
40	16.4	74.5	804	10	AL041092
41	16.4	74.5	833	10	BE412370
42	16.2	73.6	122	11	BG989620
43	16.2	73.6	400	13	AQ852470
44	16.2	73.6	406	13	AZ160523
45	16.2	73.6	431	10	AV655742

ALIGNMENTS

RESULT 1

BE377681 LOCUS 601229824F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3594076 5', mRNA EST 21-JUL-2000  
DEFINITION mRNA sequence.  
ACCESSION BE377681  
VERSION BE377681.1 GI:9323046  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 653)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LIA88767 row: o column: 05  
High quality sequence stop: 612.  
Location/Qualifiers  
1. .653  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3594076"  
/clone\_lib="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"

/dev\_stage="3 months, virgin"  
/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 132 a 181 c 195 g 145 t  
ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 653;  
Best Local Similarity 90.5%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgac 21  
||||| ||||| |||||  
Db 193 GCCAGTCTGTGAGCTCGAC 213

## RESULT 2

AI561687 244 bp mRNA EST 25-MAR-1999  
LOCUS  
DEFINITION vw88f09.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:1262057 3', mRNA sequence.  
ACCESSION AI561687  
VERSION  
KEYWORDS  
SOURCE EST.  
house mouse.  
Mus musculus

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 244)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person  
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter  
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R., and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)

TITLE  
JOURNAL  
COMMENT  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:664609  
This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end  
High quality sequence stop: 229.

FEATURES  
source

1..244  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1262057"  
/clone\_lib="Stratagene mouse skin (#937313)"  
/sex="females"  
/tissue\_type="whole skin"  
/dev\_stage="11 weeks old"  
/lab\_host="SOLR (kanamycin resistant)"  
; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dt. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'  
53 a 53 c 51 g 87 t

BASE COUNT  
ORIGIN

Query Match 76.4%; Score 16.8; DB 10; Length 244;

Best Local Similarity 90.0%; Pred. No. 5.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 cagttctgtgaacatcgacc 22  
||||| ||||| |||||  
Db 85 CAGTTCCTTTGAACATCTACC 104

## RESULT 3

AZ777264 279 bp DNA GSS 16-FEB-2001  
LOCUS  
DEFINITION 2M0011K06R Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
clone UUGC2M0011K06 R, DNA sequence.

ACCESSION AZ777264  
VERSION  
KEYWORDS  
SOURCE GSS.  
house mouse.  
Mus musculus

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 279)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.  
and Wright, D. Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)

JOURNAL  
COMMENT

Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0011 row: K column: 06  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 279.  
Location/Qualifiers  
1..279  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0011K06"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gii4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

FEATURES  
source

1..279  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0011K06"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gii4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

80 a 65 c 69 g 65 t

```

Query Match      76.4%; Score 16.8; DB 13; Length 279;
Best Local Similarity 90.0%; Pred. No. 5.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcga 20
| | | | | | | | | | | | | | | |
Db 162 CTCAGTTCTGTGACCATCGA 181

RESULT 4
LOCUS AI019782 346 bp mRNA EST 16-JUN-1998
DEFINITION ua90h12.rl Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:1364807 5', mRNA sequence.
ACCESSION AI019782
VERSION AI019782.1 GI:3234118
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 346)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:898027
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 285.
FEATURES
Location/Qualifiers
1..346
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1364807"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTTCACCAATCTGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 74 a 95 c 120 g 57 t
ORIGIN

Query Match      76.4%; Score 16.8; DB 10; Length 346;
Best Local Similarity 90.0%; Pred. No. 5.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cgcagttctgtgaacatcgac 21
| | | | | | | | | | | | | | | |
Db 162 CTCAGTTCTGTGACCATCGA 181

RESULT 5
LOCUS FR0030063 605 bp DNA GSS 25-JUN-1998
DEFINITION Fugu rubripes GSS sequence, clone 072H16aE4, genomic survey
sequence.
ACCESSION AL026432
VERSION AL026432.1 GI:3263775
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes
REFERENCE 1 (bases 1 to 605)
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,
Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:
biohelp@hgm.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
FEATURES
Location/Qualifiers
1..605
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 072H16"
/clone="072H16aE4"
BASE COUNT 139 a 147 c 173 g 132 t 14 others
ORIGIN

Query Match      76.4%; Score 16.8; DB 13; Length 605;
Best Local Similarity 90.0%; Pred. No. 6.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcga 20
| | | | | | | | | | | | | | | |
Db 29 CCCAGTTTGTGACCATCGA 48

RESULT 6
LOCUS FR0030058 611 bp DNA GSS 25-JUN-1998
DEFINITION Fugu rubripes GSS sequence, clone 072H16aG6, genomic survey
sequence.
ACCESSION AL026427
VERSION AL026427.1 GI:3263770
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes
REFERENCE 1 (bases 1 to 611)
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y.,
Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource
Centre, Hinxton, Cambridge, CB10 1SB, UK. Email:
biohelp@hgm.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS

```

DESCR:  
One pass dye-terminator sequencing of cosmid cloned genomic sequence.

#### FEATURES

Location/Qualifiers  
1..611  
/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid 072H16"  
/clone="072H16G6"

BASE COUNT 154 a 144 c 169 g 128 t 16 others  
ORIGIN

Query Match 76.4%; Score 16.8; DB 13; Length 611;

Best Local Similarity 90.0%; Pred. NO. 6.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccagttctgtgaacatcga 20  
||||| ||||||| |||||||

Db 19 CCCAGTTTGTGGAACATCGA 38

#### RESULT 7

AZ777336 613 bp DNA 16-FEB-2001  
LOCUS 2M0011108R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC2M0011108 R, DNA sequence.

ACCESSION AZ777336

VERSION AZ777336.1 GI:12905813

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 613)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert length: 10000 Std Error: 0.00

Plate: 0011 row: 1 column: 08

Seq primer: CACACAGGAACACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 613.

Location/Qualifiers

1..613

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0011108"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/notes="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 174 a 121 c 142 g 176 t  
ORIGIN

Query Match 76.4%; Score 16.8; DB 13; Length 613;

Best Local Similarity 90.0%; Pred. NO. 6.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccagttctgtgaacatcga 20  
||||| ||||||| |||||||

Db 495 CTCAGTTCTGTGAATATCGA 514

#### RESULT 8

AK006515

LOCUS

DEFINITION

AK006515 1301 bp mRNA HTC 05-JUL-2001

Mus musculus adult male testis cDNA, RIKEN full-length enriched

library, clone:1700029008, full insert sequence.

ACCESSION AK006515

VERSION AK006515.1 GI:12839656

KEYWORDS CAP trapper.

SOURCE Mus musculus

clone:1700029008.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1301)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Methods in enzymology. 303, 19-44 (1999)

99279253

PUBLISHED 10349636

REFERENCE 2 (bases 1 to 1301)

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome research. 10 (10), 1617-1630 (2000)

20499374

PUBLISHED 11042159

REFERENCE 3 (bases 1 to 1301)

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multichannel sequencer

Genome research. 10 (11), 1757-1771 (2000)

20530913

PUBLISHED 11076861

REFERENCE 4 (bases 1 to 1301)

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 1301)

ADACHI, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,

Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,

Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F.,



**TITLE**  
JOURNAL

Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

**Direct Submission**  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT**  
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGAGCGCGCAATTAATTCGAGTTAATTAATTAATCCCGCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

# FEATURES

**source**  
Location/Qualifiers  
1. .206  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2590316"  
/clone\_lib="NCI\_CGAP\_Ov32"  
/sex="female"  
/tissue\_type="papillary serous carcinoma"  
/lab\_host="DH10B"  
/note="Organ: ovary; Vector: pAMPl; mRNA made from ovarian carcinoma. cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 500 bp. Non-amplified library."

**BASE COUNT**  
64 a 24 c 24 g 94 t

**ORIGIN**

**Query Match**  
Best Local Similarity 74.5%; Score 16.4; DB 10; Length 206;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**Oy** 4 agttctgtgaacatgcac 21  
|||||  
**Db** 35 AGTCTGTGACATGTGAC 52

**RESULT** 10  
**LOCUS** AA701011 234 bp mRNA EST 19-DEC-1997  
**DEFINITION** ZG55d04.sl Soares\_pineal\_gland\_N3HPG Homo sapiens cDNA clone IMAGE:397255 3', mRNA sequence.  
**ACCESSION** AA701011  
**VERSION** AA701011.1 GI:2704176  
**KEYWORDS** EST.  
**SOURCE** human.

**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE**  
**AUTHORS** 1 (bases 1 to 234)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Willson, R.  
**TITLE** WashU-NCI human EST Project  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Willson RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

# FEATURES

**source**  
Location/Qualifiers  
1. .1301  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGD:MGI:1902349"  
/db\_xref="MGD:MGI:1917320"  
/clone="1700029008"  
/sex="male"  
/tissue\_type="testis"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
75. .1058  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB24628.1"  
/db\_xref="GI:12839657"  
/translation="MGFGSSSAGPNLKEWLREQFDHPLEHCDLTRHDAAYVDLQ  
TLRLQESYRSRINSGKSWCCWLEPTPLRIATAAGHNCVDFLRKGAEDVLVDY  
KGOTALYVAVVNGHLESTEILLEAGADPNGRHHRSTPVYHASRVGRDDILKALIRYG  
ADVDVNHLLTPDTPPPSRRLTSLVVCPLYLSAAYHNLCQFRLLQAGNPDFNCNGP  
VNTQEFYRGSPGVMDAVLRHGCEAFVSLIVERGANLNLVKWESLPEARGRKMDP  
ENLQVFKEARSIPRTLISLCRVAVRRAALGYRLHLVPAARPHKEVFAL"

**BASE COUNT** 274 a 345 c 368 g 313 t 1 others

**ORIGIN**

**Query Match**  
Best Local Similarity 76.4%; Score 16.8; DB 12; Length 1301;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**Oy** 2 gcagttctgtgaacatgcac 21  
|||||  
**Db** 131 GCAGTCTGTGACCATCCAC 150

**RESULT** 9  
**LOCUS** AW086177 206 bp mRNA EST 14-OCT-1999  
**DEFINITION** xc77ell.x1 NCI\_CGAP\_Ov32 Homo sapiens cDNA clone IMAGE:2590316 3', mRNA sequence.  
**ACCESSION** AW086177  
**VERSION** AW086177.1 GI:6041251

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham.

## FEATURES

## source

## Location/Qualifiers

1..234

/organism="Homo sapiens"

/db\_xref="GDB:1303157"

/db\_xref="taxon:9606"

/clone="IMAGE:397255"

/clone.lib="Soares\_pineal\_gland\_N3HPC"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: pineal gland; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCAATCTGAAGTGGAGCGCGCTTTTCTTTTCTTTT 3'] ; double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

79 a 31 c 35 g 89 t

## BASE COUNT

## ORIGIN

## Query Match

Best Local Similarity 74.5%; Score 16.4; DB 10; Length 234;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21

|||||

Db 21 AGTTCTGTGAACATGTGAC 38

## RESULT 11

## F02043

## LOCUS

DEFINITION F02043 270 bp mRNA EST 01-FEB-1995  
HSCOMG102 normalized infant brain cDNA Homo sapiens cDNA clone

c-0mg10 3', mRNA sequence.

ACCESSION F02043.1 GI:645600

VERSION F02043.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Auffray,C., Behar,G., Bols,F., Bouchier,C., da Silva,C., Devignes

,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,

Mitchell,H., Maritage-Sanson,R., Pietsch,G., Pouliot,Y.,

Sebastiani-Kabaktchi,S.C. and Tessier,A.

IMAGE: molecular integration of the analysis of the human genome

and its expression

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

JOURNAL 95277534

MEDLINE

COMMENT

Contact: Genethon

Genethon Centre de recherche sur le Genome Humain

1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE

Tel: 33169472800

Fax: 33160778698

Email: genexpres@genethon.fr

Single read, removed at sequence 5'end

Genexpres\_library\_idt: C; Genexpres\_sequence\_idt: alc-0mg10

Seq primer: (-21)M13-universal.

Location/Qualifiers

1..270

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="c-0mg10"

/clone.lib="normalized infant brain cDNA"

/sex="Female"

/tissue\_type="total brain"

/dev\_stage="3 months old"

## FEATURES

## source

## Location/Qualifiers

1..270

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="c-0mg10"

/clone.lib="normalized infant brain cDNA"

/sex="Female"

/tissue\_type="total brain"

/dev\_stage="3 months old"

/note="Organ: brain; Vector: lafmid BA; Site\_1: HindIII; Site\_2: NotI; sex=Female; dev\_stage=3 months old; isolate=muscular atrophy patient; tissue\_type=total brain ; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Soaress, psychiatry Dept. Columbia University, USA. Normalization\_method: Bento Soares, P.N.A.S in press"

BASE COUNT 86 a 38 c 39 g 107 t

## ORIGIN

## Query Match

Best Local Similarity 74.5%; Score 16.4; DB 11; Length 270;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21

|||||

Db 23 AGTTCTGTGAACATGTGAC 40

## RESULT 12

## AA569516

## LOCUS

DEFINITION nf23e01.s1 NCI-CGAP\_Prl Homo sapiens cDNA clone IMAGE:914616, mRNA sequence.

ACCESSION AA569516

VERSION AA569516.1 GI:2343496

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 290)

AUTHORS NCI-CGAP

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.

, Michael Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

DNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.

Clone distribution: Washington University Genome Sequencing Center

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 263.

Location/Qualifiers

1..290

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:914616"

/clone.lib="NCI-CGAP\_Prl"

/sex="Male"

/dev\_stage="45 years old"

/lab\_host="DH10B"

/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st

strand cDNA was primed with oligo(dT)17 on 50 ng of

DNase-treated, total cellular RNA obtained from 5,000-10

,000 microdissected, histologically normal prostate

epithelial cells. Double-stranded cDNA was ligated to

EcoRI adaptors, 5 cycles of PCR applied to the cDNA with

an adaptor-specific primer, and the resulting PCR product

subcloned into pAMP10 by the UDG-cloning method (Life

Technologies). Average insert size is 600 bp. NOTE: Not

directionally cloned. This library was constructed by

David Krizman."

77 a 76 c 73 g 64 t

## BASE COUNT

## ORIGIN

Query Match 74.5%; Score 16.4; DB 10; Length 290;  
 Best Local Similarity 94.4%; Pred. No. 8.6e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 gttctgtgaacatgcacc 22  
 ||||| ||||| ||||| |||||  
 Db 233 GTTCTGTGGACATGCACC 250

RESULT 13  
 F03921

LOCUS  
 DEFINITION HSC2DF112 normalized infant brain cDNA Homo sapiens CDNA clone  
 C-2df11 3', mRNA sequence.

ACCESSION F03921  
 VERSION F03921  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 303)  
 AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Degniges  
 , M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F.,  
 Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,  
 Sebastiani-Kabatchis, C. and Tessier, A.  
 IMAGE: molecular integration of the analysis of the human genome  
 and its expression  
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 95277534  
 MEDLINE  
 COMMENT Contact: Genethon  
 Genethon  
 Genethon Centre de recherche sur le Genome Humain  
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 33160778698  
 Email: genexpress@genethon.fr  
 Single read. removed at sequence 5' end  
 Genexpress\_library\_id: C; Genexpress\_sequence\_id: alc-2df11  
 Seq primer: (-21)M13\_universal.

FEATURES  
 source Location/Qualifiers  
 1..303  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="c-2df11"  
 /clone\_lib="normalized infant brain cDNA"  
 /sex="Female"  
 /tissue\_type="total brain"  
 /dev\_stage="3 months old"  
 /note="Organ: brain; Vector: lafmid BA; Site\_1: HindIII;  
 Site\_2: NotI; sex=Female; dev\_stage=3 months old;  
 isolate=muscular atrophy patient; tissue\_type=total brain  
 ; total mRNA was oligo-(dT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the  
 lafmid BA vector. Clone library from B.Souares, Psychiatry  
 Dept. Columbia University, USA. Normalization\_method:  
 Bento Soares, P.N.A.S in press"  
 99 a 42 c 43 g 116 t 3 others

BASE COUNT  
 ORIGIN

Query Match 74.5%; Score 16.4; DB 11; Length 303;  
 Best Local Similarity 94.4%; Pred. No. 8.6e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatgcac 21  
 ||||| ||||| ||||| |||||  
 Db 23 AGTTCGTGAACATTGAC 40

RESULT 14  
 AA843570

LOCUS  
 DEFINITION aj54f09.s1 Soares\_testis\_NHT Homo sapiens CDNA clone IMAGE:1394153  
 3', mRNA sequence.

ACCESSION AA843570  
 VERSION AA843570.1 GI:2930088  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 304)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb-remail.nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Ronaldo  
 , Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 1108 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 284.

FEATURES  
 source Location/Qualifiers  
 1..304  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1394153"  
 /clone\_lib="Soares\_testis\_NHT"  
 /sex="Male"  
 /lab\_host="DH10B"  
 /note="vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech Laboratories  
 , Inc., and primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization to Cots, and was  
 constructed by Bento Soares and M. Fatima Ronaldo. "  
 99 a 43 c 43 g 119 t

BASE COUNT  
 ORIGIN

Query Match 74.5%; Score 16.4; DB 10; Length 304;  
 Best Local Similarity 94.4%; Pred. No. 8.7e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatgcac 21  
 ||||| ||||| ||||| |||||  
 Db 26 AGTTCGTGAACATTGAC 43

RESULT 15  
 AI084952

LOCUS  
 DEFINITION ow86c05.s1 Soares\_fetal\_liver\_spleen\_lNFLS\_s1 Homo sapiens CDNA  
 clone IMAGE:1653704 3', mRNA sequence.

ACCESSION AI084952  
 VERSION AI084952.1 GI:3423375  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 335)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Search completed: February 25, 2002, 17:21:24  
Job time: 16177 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:53:12 ; Search time 2331.3 Seconds  
(without alignments)  
148.604 Million cell updates/sec

Title: US-09-698-903B-2

Perfect score: 21

Sequence: 1 gtaacatagatgacaccgcgc 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 824859755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_em.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	6	AX127749 Sequence
C 2	21	100.0	249	6	AR110601 Sequence
C 3	21	100.0	249	6	AR151000 Sequence
C 4	21	100.0	249	6	AR152423 Sequence
C 5	21	100.0	249	6	AR152432 Sequence
C 6	21	100.0	261	6	AR110597 Sequence
C 7	21	100.0	261	6	AR150996 Sequence
C 8	21	100.0	261	6	AR152419 Sequence
C 9	21	100.0	261	6	AR152428 Sequence
C 10	21	100.0	280	6	E01312
C 11	21	100.0	288	11	RIC1429B
C 12	21	100.0	318	1	ABA7623
C 13	21	100.0	358	1	ABA7624
C 14	21	100.0	563	6	A71436 Sequence 6
C 15	21	100.0	838	6	AX014764 Sequence
C 16	21	100.0	900	6	AX019188 Sequence
C 17	21	100.0	900	6	AX019212 Sequence
C 18	21	100.0	1036	6	AX014765 Sequence
C 19	21	100.0	1045	12	SCO308515
C 20	21	100.0	1138	6	A59870
C 21	21	100.0	1138	6	A94795 Sequence 39
C 22	21	100.0	1287	6	A18052
C 23	21	100.0	1287	6	AR095108 Sequence
C 24	21	100.0	1287	6	AR098314 Sequence
C 25	21	100.0	1287	6	AX012339 Sequence
C 26	21	100.0	1287	6	I49887 Sequence 3
C 27	21	100.0	1287	6	I82375 Sequence 3
C 28	21	100.0	1559	6	A84754 Sequence 6
C 29	21	100.0	1559	6	AR118432 Sequence
C 30	21	100.0	1690	6	AX021176 Sequence
C 31	21	100.0	1829	6	A26453 Recombinant
C 32	21	100.0	1829	6	I16753 Sequence 17
C 33	21	100.0	1863	6	A32909 Chimeric en
C 34	21	100.0	1863	6	A40108 Sequence 16
C 35	21	100.0	1863	6	AR022666 Sequence
C 36	21	100.0	1863	6	AR029592 Sequence
C 37	21	100.0	1949	6	A83076 Sequence 10
C 38	21	100.0	2020	6	D88395 Oryza sativ
C 39	21	100.0	2042	12	AF190131 Cloning v
C 40	21	100.0	2115	6	AR027732 Sequence
C 41	21	100.0	2145	6	AR019221 Sequence
C 42	21	100.0	2256	8	AB003140 Arabidops
C 43	21	100.0	2319	6	A84753 Sequence 5
C 44	21	100.0	2319	6	AR118431 Sequence
C 45	21	100.0	2345	6	AR118422 Sequence

ALIGNMENTS

RESULT 1	AX127749	AX127749	21 bp	DNA	PAT	15-MAY-2001
LOCUS	Sequence 2 from Patent WO0131042.					
DEFINITION	AX127749					
ACCESSION	AX127749					
VERSION	AX127749.1	GI:14134396				
KEYWORDS	synthetic construct.					
SOURCE	artificial construct					
ORGANISM	1 (bases 1 to 21)					
REFERENCE	Weston,B. and de Beuckeleer,M.					
AUTHORS	Male-sterile brassica plants and methods for producing same					
TITLE	Patent: WO 0131042-A 2 03-MAY-2001;					
JOURNAL	Aventis CropScience N.V. (BE)					
FEATURES	Location/Qualifiers					
source	1. .21					
	/organism="synthetic construct"					
	/db_xref="taxon:32630"					
	/note="primer MDB355"					

BASE COUNT 7 a 6 c 5 g 3 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21  
|||||

Db 1 GTAACATAGATGACACCGCGC 21

## RESULT 2

AR110601/c  
LOCUS AR110601 249 bp DNA PAT 14-FEB-2001  
DEFINITION Sequence 8 from patent US 6114608.  
ACCESSION AR110601  
VERSION AR110601.1 GI:12826877  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 249)  
AUTHORS Mettler,I.J., Dietrich,P.S and Sinibaldi,R.M.  
TITLE Nucleic acid construct comprising bacillus thuringiensis cryIAb gene  
JOURNAL Patent: US 6114608-A 8 05-SEP-2000;  
FEATURES Location/Qualifiers  
source 1..249

BASE COUNT 79 a 38 c 47 g 85 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 3

AR151000/c  
LOCUS AR151000 249 bp DNA PAT 08-AUG-2001  
DEFINITION Sequence 8 from patent US 6229075.  
ACCESSION AR151000  
VERSION AR151000.1 GI:15115591  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 249)  
AUTHORS Mettler,I.J., Plaisted,D.C., Grier,S.L., Houghton,W. and Gardiner,M.  
TITLE Inbred maize line R412H  
JOURNAL Patent: US 6229075-A 8 08-MAY-2001;  
FEATURES Location/Qualifiers  
source 1..249

BASE COUNT 79 a 38 c 47 g 85 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 4

AR152423/c  
LOCUS AR152423 249 bp DNA PAT 08-AUG-2001  
DEFINITION Sequence 8 from patent US 6232533.  
ACCESSION AR152423  
VERSION AR152423.1 GI:15118473  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 249)  
AUTHORS Mettler,I.J., Plaisted,D.C., Grier,S.L., Houghton,W. and Gardiner,M.  
TITLE Inbred maize line R372H  
JOURNAL Patent: US 6232533-A 8 15-MAY-2001;  
FEATURES Location/Qualifiers  
source 1..249

BASE COUNT 79 a 38 c 47 g 85 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 5

AR152432/c  
LOCUS AR152432 249 bp DNA PAT 08-AUG-2001  
DEFINITION Sequence 8 from patent US 6232534.  
ACCESSION AR152432  
VERSION AR152432.1 GI:15118482  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 249)  
AUTHORS Mettler,I.J., Plaisted,D.C., Grier,S.L., Houghton,W. and Gardiner,M.  
TITLE Inbred maize line R660H  
JOURNAL Patent: US 6232534-A 8 15-MAY-2001;  
FEATURES Location/Qualifiers  
source 1..249

BASE COUNT 79 a 38 c 47 g 85 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 6

AR110597/c  
LOCUS AR110597 261 bp DNA PAT 14-FEB-2001  
DEFINITION Sequence 4 from patent US 6114608.  
ACCESSION AR110597  
VERSION AR110597.1 GI:12826873  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

Unclassified.  
1 (bases 1 to 261)  
REFERENCE Mettler, I.J., Dietrich, P.S and Sinibaldi, R.M.  
AUTHORS Nucleic acid construct comprising bacillus thuringiensis cryIab  
TITLE gene  
JOURNAL Patent: US 6114608-A 4 05-SEP-2000;  
FEATURES Location/Qualifiers  
          1..261  
          /organism="unknown"  
BASE COUNT 83 a 41 c 50 g 87 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 7  
ARI50996/c  
LOCUS ARI50996 261 bp DNA 08-AUG-2001  
DEFINITION Sequence 4 from patent US 6229075.  
ACCESSION ARI50996  
VERSION ARI50996.1 GI:15115587  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 261)  
AUTHORS Mettler, I.J., Plaisted, D.C., Grier, S.L., Houghton, W. and  
Gardiner, M.  
TITLE Inbred maize line R412H  
JOURNAL Patent: US 6229075-A 4 08-MAY-2001;  
FEATURES Location/Qualifiers  
          1..261  
          /organism="unknown"  
BASE COUNT 83 a 41 c 50 g 87 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 8  
ARI52419/c  
LOCUS ARI52419 261 bp DNA 08-AUG-2001  
DEFINITION Sequence 4 from patent US 6232533.  
ACCESSION ARI52419  
VERSION ARI52419.1 GI:15118469  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 261)  
AUTHORS Mettler, I.J., Plaisted, D.C., Grier, S.L., Houghton, W. and  
Gardiner, M.  
TITLE Inbred maize line R372H  
JOURNAL Patent: US 6232533-A 4 15-MAY-2001;  
FEATURES Location/Qualifiers  
          1..261  
          /organism="unknown"  
BASE COUNT 83 a 41 c 50 g 87 t  
ORIGIN

Unclassified.  
1 (bases 1 to 261)  
REFERENCE Mettler, I.J., Dietrich, P.S and Sinibaldi, R.M.  
AUTHORS Nucleic acid construct comprising bacillus thuringiensis cryIab  
TITLE gene  
JOURNAL Patent: US 6114608-A 4 05-SEP-2000;  
FEATURES Location/Qualifiers  
          1..261  
          /organism="unknown"  
BASE COUNT 83 a 41 c 50 g 87 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 9  
ARI52428/c  
LOCUS ARI52428 261 bp DNA 08-AUG-2001  
DEFINITION Sequence 4 from patent US 6232534.  
ACCESSION ARI52428  
VERSION ARI52428.1 GI:15118478  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 261)  
AUTHORS Mettler, I.J., Plaisted, D.C., Grier, S.L., Houghton, W. and  
Gardiner, M.  
TITLE Inbred maize line R660H  
JOURNAL Patent: US 6232534-A 4 15-MAY-2001;  
FEATURES Location/Qualifiers  
          1..261  
          /organism="unknown"  
BASE COUNT 83 a 41 c 50 g 87 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 10  
E01312/c  
LOCUS E01312 280 bp DNA 29-SEP-1997  
DEFINITION 3' Untranslated region of nopalín synthetase gene.  
ACCESSION E01312  
VERSION E01312.1 GI:2169571  
KEYWORDS JP 1987201527-A/4.  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 280)  
AUTHORS Rojlyaa, E.B., Robaato, T.F. and Suteilibun, J.R.  
TITLE PROTECTION OF PLANT FROM VIRUS INFECTION  
JOURNAL Patent: JP 1987201527-A 4 05-SEP-1987;  
COMMENT MONSANTO CO, WASHINGTON UNIV  
OS Agrobacterium  
PN JP 1987201527-A/4  
PD 05-SEP-1987  
PF 29-OCT-1986 JP 1986258063  
PR 29-OCT-1985 US 85 792389, 27-MAR-1986 US 86 844918, PR  
09-OCT-1986 US 86 917027  
PI ROJLYAA, ENU BIICHI, ROBAATO, TEII FURAREI,  
PI SUTEILIBUN, JII ROJLYAASU  
PC AOIHS/00, AOIHI/00, C12N1/20, C12N15/00, C07K13/00, PC  
(C12N1/20,  
PC C12R1:01);  
CC strandedness: Single;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
FH Key Location/Qualifiers

```

FH
FT polyA_signal >1..<281.
  Location/Qualifiers
    1..280
      /organism="unidentified"
      /db_xref="taxon:32644"
BASE COUNT      83 a 49 c 58 g 90 t
ORIGIN

Query Match      100.0%; Score 21; DB 6; Length 280;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
    |||
Db 246 GTAACATAGATGACACCGCGC 236

RESULT 11
RICL429B/c
LOCUS
DEFINITION Rice genomic DNA, L429B, sequence tagged site.
ACCESSION D25449
VERSION D25449.1 GI:436695
KEYWORDS STS; Not I linking clone; RFLP; Rice.
SOURCE Oryza sativa (strain:Nipponbare) DNA.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 288)
Minobe, Y.
Direct Submission
Submitted (18-NOV-1993) to the DDBJ/EMBL/GenBank databases. Yuzo
Minobe, National Institute of Agrobiological Resources, Rice Genome
Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305, Japan
(E-mail:MINOBE@rcs01.riken.go.jp, Tel:0298-38-7441,
Fax:0298-38-7468)
2 (bases 1 to 288)
Minobe, Y.
Nucleotide sequence of sequence tagged site from japonica rice
Nipponbare as an RFLP marker
Unpublished (1993)
Submitted (18-Nov-1993) to DDBJ by:
Yuzo Minobe
Dept. Rice Genome Research Program
National Institute of Agrobiological Resources
Kannondai 2-1-2
Tsukuba, Ibaraki
Japan
Phone: 0298-38-7441
Fax: 0298-38-7468
PROJECT "RGP".
  Location/Qualifiers
    1..288
      /organism="Oryza sativa"
      /strain="Nipponbare"
      /db_xref="taxon:4530"
BASE COUNT      77 a 61 c 80 g 68 t 2 others
ORIGIN

Query Match      100.0%; Score 21; DB 11; Length 288;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
    |||
Db 113 GTAACATAGATGACACCGCGC 93

RESULT 12

```

---

```

ABA7623
LOCUS
DEFINITION ABA7623 318 bp DNA BCT 18-FEB-2000
Agrobacterium tumefaciens nos terminator with artificial insert of
40 bp.
ACCESSION AJ007623
VERSION AJ007623.1 GI:3319860
KEYWORDS nos gene; terminator
SOURCE Agrobacterium tumefaciens.
ORGANISM Agrobacterium tumefaciens.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
1 (bases 1 to 318)
Hardegger, M.R., Brodmann, P. and Herrmann, A.
Quantitative detection of the 35S promoter and the NOS terminator
using quantitative competitive PCR
Z. Lebensm.-Unters. -Forsch., A Eur. Food Res. Technol. 209, 83-87
(1999)
2 (bases 1 to 318)
Hardegger, M.R.
Direct Submission
Submitted (03-JUL-1998) Hardegger M.R., Lebensmittel, Kantonales
Laboratorium Basel-Stadt, Kannenfeldstrasse 2, CH/4012 Basel,
SWITZERLAND
  Location/Qualifiers
    1..318
      /organism="Agrobacterium tumefaciens"
      /plasmid="Ti"
      /db_xref="taxon:358"
    1..318
      /gene="nos"
    1..318
      /gene="nos"
      /note="insert location 142-181"
BASE COUNT      102 a 64 c 55 g 97 t
ORIGIN

Query Match      100.0%; Score 21; DB 1; Length 318;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
    |||
Db 16 GTAACATAGATGACACCGCGC 36

RESULT 13
ABA7624
LOCUS
DEFINITION ABA7624 358 bp DNA BCT 18-FEB-2000
Agrobacterium tumefaciens nos terminator with artificial insert of
80 bp.
ACCESSION AJ007624
VERSION AJ007624.1 GI:3319861
KEYWORDS nos gene; terminator
SOURCE Agrobacterium tumefaciens.
ORGANISM Agrobacterium tumefaciens.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
1 (bases 1 to 358)
Hardegger, M.R., Brodmann, P. and Herrmann, A.
Quantitative detection of the 35S promoter and the NOS terminator
using quantitative competitive PCR
Z. Lebensm.-Unters. -Forsch., A Eur. Food Res. Technol. 209, 83-87
(1999)
2 (bases 1 to 358)
Hardegger, M.R.
Direct Submission
Submitted (03-JUL-1998) Hardegger M.R., Lebensmittel, Kantonales
Laboratorium Basel-Stadt, Kannenfeldstrasse 2, CH/4012 Basel,
SWITZERLAND
  Location/Qualifiers
    1..358
      /organism="Agrobacterium tumefaciens"

```



```

misc_structure 533. .568
terminator    569. .832
BASE COUNT   296 a    175 g    233 t
ORIGIN
Query Match          100.0%; Score 21; DB 6; Length 838;
Best Local Similarity 100.0%; Pred. NO. 0.26;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 gtaacatagatgacacccgc 21
        | | | | | | | | | | | | | |
Db      822 GTAACATAGATGACACCCGC 802

Search completed: February 25, 2002, 17:59:54
Job time: 18402 sec

```

gene	/plasmid="Ti"																							
terminator	1. .358																							
	/gene="nos"																							
	1. .358																							
	/gene="nos"																							
BASE COUNT	114	a	75	c	63	g	106	t																
ORIGIN																								
Query Match 100.0%; Score 21; DB 1; Length 358;																								
Best Local Similarity	100.0%;	Pred. No. 0.23;																						
Matches	21;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;															
Qy	1	gtaacatagatgacacgcgcg 21																						
Db	16	GTAACATAGATGACACCGCGC 36																						
RESULT 14																								
A71436/c																								
LOCUS	A71436	563	bp	DNA																				
DEFINITION	Sequence 6 from Patent WO9810081.				PAT	07-MAY-1999																		
ACCESSION	A71436																							
VERSION	A71436.1	GI:4775049																						
KEYWORDS	.																							
SOURCE	unidentified.																							
ORGANISM	unidentified.																							
REFERENCE	1. (bases 1 to 563)																							
AUTHORS	Michiels,F. and Williams,M.																							
TITLE	IMPROVED BARSTAR GENE																							
JOURNAL	Patent: WO 9810081-A 6 12-MAR-1998;																							
FEATURES	MICHIELS FRANK (BE)																							
source	Location/Qualifiers																							
	1.	.563																						
BASE COUNT	152	a	133	c	145	g	133	t																
ORIGIN																								
Query Match 100.0%; Score 21; DB 6; Length 563;																								
Best Local Similarity	100.0%;	Pred. No. 0.25;																						
Matches	21;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;															
Qy	1	gtaacatagatgacacgcgcg 21																						
Db	538	GTAACATAGATGACACCGCGC 518																						
RESULT 15																								
AX014764/c																								
LOCUS	AX014764	838	bp	DNA																				

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 13:39:22 ; Search time 716.55 Seconds  
(without alignments)  
25.126 Million cell updates/sec

Title: US-09-698-903B-2

Perfect score: 21  
Sequence: 1 gtaacatagatgacacggc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	22	PCR primer MDB355
C 2	21	100.0	249	21	Nucleotide sequenc
C 3	21	100.0	249	22	Nopline synthetas
C 4	21	100.0	249	22	3' nopline synth
C 5	21	100.0	249	22	Nucleotide sequenc
C 6	21	100.0	249	22	Nucleotide sequenc
C 7	21	100.0	253	21	3'UTR of nopline
C 8	21	100.0	254	21	Nos terminator nuc
C 9	21	100.0	257	21	Agrobacterium nos
C 10	21	100.0	261	21	Nucleotide sequenc
C 11	21	100.0	261	22	Nopline synthetas

C 12	21	100.0	261	22	3' nopline synth
C 13	21	100.0	261	22	Nucleotide sequenc
C 14	21	100.0	261	22	Nucleotide sequenc
C 15	21	100.0	263	18	Nopline synthase
C 16	21	100.0	270	20	NOS terminator seq
C 17	21	100.0	270	22	Tomato spotted wil
C 18	21	100.0	563	19	Part of plasmid pL
C 19	21	100.0	900	20	vector plasmid pSI
C 20	21	100.0	1138	18	PMJB1 DNA sequence
C 21	21	100.0	1138	18	PMJB1 used
C 22	21	100.0	1138	20	Plasmid PMJB1 EcoR
C 23	21	100.0	1559	19	Arabidopsis thalia
C 24	21	100.0	1630	21	Expression cassett
C 25	21	100.0	1690	20	Nematode-responsiv
C 26	21	100.0	1829	13	35S CamV promoter-
C 27	21	100.0	1863	13	Sequence of chim
C 28	21	100.0	1863	15	Tomato-tobacco end
C 29	21	100.0	1923	14	OBFlt sequence.
C 30	21	100.0	1949	20	Nucleotide sequenc
C 31	21	100.0	2115	19	Oleosin-hirudin fu
C 32	21	100.0	2145	18	Raspberry drul gen
C 33	21	100.0	2275	22	Oligonucleotide #1
C 34	21	100.0	2319	19	Arabidopsis thalia
C 35	21	100.0	2345	19	Nucleotide sequenc
C 36	21	100.0	2345	20	T-DNA sequence of
C 37	21	100.0	2345	22	Agrobacterium tume
C 38	21	100.0	2378	21	Expression cassett
C 39	21	100.0	2436	21	Expression cassett
C 40	21	100.0	2543	22	DNA construct codi
C 41	21	100.0	2709	19	Oleosin-protein A
C 42	21	100.0	2917	18	Raspberry drul gen
C 43	21	100.0	3069	22	pND3018 repression
C 44	21	100.0	3121	22	pND3008 activation
C 45	21	100.0	3183	17	SAG12-1 promoter,

#### ALIGNMENTS

RESULT	1
AD06991	AD06991 standard; DNA; 21 BP.
XX	AD06991;
XX	AD06991;
DT	06-AUG-2001 (first entry)
XX	PCR primer MDB355 to amplify a fragment of pTCO113.
DE	Plasmid pTCO113; transgenic Brassica plant; transformation event;
XX	male-sterility gene; tobacco; PCR primer; probe; ss.
KW	Chimeric - Bacillus amyloliquefaciens.
XX	Chimeric - Nicotiana tabacum.
OS	WO200131042-A2.
PN	03-MAY-2001.
XX	26-OCT-2000; 2000WO-EP10680.
XX	29-OCT-1999; 99US-0430497.
XX	(AVET ) AVENTIS CROPS SCIENCE NV.
PA	Weston B, De Beuckeleer M;
PI	WPI; 2001-300517/31.
XX	Transgenic Brassica plants, seeds, cells or tissues, characterized by
PT	harboring specific transformation events, particularly by presence of
PT	male-sterility gene, at specific location in its genome -
XX	

PS Claim 1; Page 26; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.

CC The present sequence is PCR primer MDB355 used to amplify a fragment of  
 CC plasmid pTCOLL3. The primer is also used as a probe. The amplified  
 CC fragment comprises coding region of barnase from  
 CC *Bacillus amyloliquefaciens* and promoter region of the anther-specific  
 CC gene TA29 from *Nicotiana tabacum*. This primer corresponds to position  
 CC 2667-2687 of plasmid pTCOLL3.

XX Sequence 21 BP; 7 A; 6 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.057;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
 |||||  
 DB 1 gtaacatagatgacacgcgc 21

RESULT 2  
 AAA75465/C  
 ID AAA75465 standard; DNA; 249 BP.  
 XX  
 AC AAA75465;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Nucleotide sequence of a nopaline synthase terminator.

KW 35S promoter; alcohol dehydrogenase intron; nopaline synthase terminator;  
 KW insecticidal Cry IAb protein toxin; transgenic plant; fungal toxin;  
 KW aflatoxin; baculovirus; Lepidopteran insect; glufosinate; ds.  
 XX Unidentified.

XX US6114608-A.  
 PN  
 XX 05-SEP-2000.  
 XX  
 XX 13-MAR-1998; 98US-0042426.  
 XX  
 XX 14-MAR-1997; 97US-0109808.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Dietrich PS, Mettler IJ, Sinibaldi RW;  
 XX  
 XX WPI; 2000-586487/55.

XX Novel DNA construct comprising an expression cassette having a  
 PT functional constitutive promoter, operably linked to a maize alcohol  
 PT dehydrogenase intron, gene encoding Cry IAb protein and a terminator -  
 XX  
 PS Claim 2; Column 25-26; 24pp; English.

XX The present sequence represents a nopaline synthase terminator, which is  
 CC used to create the construct of the invention. The specification  
 CC describes a nucleic acid construct, comprising an expression cassette  
 CC containing a cauliflower mosaic virus (CaMV) 35S promoter, a maize  
 CC alcohol dehydrogenase intron, a DNA molecule encoding an insecticidal  
 CC Cry IAb protein toxin and a NOS (nopaline synthase) terminator in  
 CC operable linkage. The nucleic acid constructs are useful for  
 CC producing transgenic plants such as maize, wheat, barley, sorghum and  
 CC rice. Preferably, maize plants such as field corn, sweet corn, white

CC corn, silage corn or popcorn are produced. The transgenic maize is  
 CC used for preparing food materials with reduced levels of fungal toxins,  
 CC e.g. aflatoxins. The transgenic plants express a baculovirus crystal  
 CC protein toxin, which is effective against Lepidopteran insects at  
 CC relatively high levels and further provides resistance to the  
 CC non-selective herbicide glufosinate.

XX Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 249;  
 Best Local Similarity 100.0%; Pred. No. 0.085;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
 |||||  
 DB 247 GTAACATAGATGACACGC GC 227

RESULT 3  
 AAD06313/C  
 ID AAD06313 standard; DNA; 249 BP.  
 XX  
 AC AAD06313;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE Nopaline synthase terminator of Pat gene expression cassette.

XX Transgenic maize; expression cassette; 35S promoter; nopaline synthetase;  
 KW NOS terminator; alcohol dehydrogenase intron; CryIAb toxin;  
 KW foodstuffs preparation; phosphinothricin acetyl transferase; PAT; ds.  
 XX Unidentified.

XX US6222104-B1.  
 XX  
 XX 24-APR-2001.  
 XX 13-APR-1999; 99US-0291238.  
 XX  
 XX 09-NOV-1994; 94US-0336627.  
 XX 22-AUG-1996; 96US-0716836.  
 XX 14-MAR-1997; 97US-0818573.  
 XX 13-MAR-1998; 98US-0042426.  
 XX (NOVS ) NOVARTIS AG.

XX Mettler IJ, Krier M, Miles D;  
 XX  
 XX WPI; 2001-327266/34.  
 XX  
 XX Novel transgenic maize seed for hybrid maize plant production,  
 PT comprising expression cassette linked operably with CaMV 35S promoter,  
 PT alcohol dehydrogenase intron, insecticidal CryIAb toxin gene or NOS  
 PT terminator -  
 XX  
 PS Claim 3; Column 25-26; 24pp; English.

XX The present invention relates to transgenic maize seed for hybrid maize  
 CC plant production, comprising an expression cassette linked operably with  
 CC constitutive Cauliflower mosaic virus (CaMV) 35S promoter, an alcohol  
 CC dehydrogenase intron, a DNA encoding an insecticidal  
 CC *Bacillus thuringiensis* (Bt) CryIAb toxin and a nopaline synthetase (NOS)  
 CC terminator. The transgenic maize seed is used in maize cultivation  
 CC methods for propagating hybrid maize seeds and for growing hybrid maize  
 CC plants. The maize is also used in foodstuffs preparation for animal or  
 CC human consumption. The inbred maize line NP948 of the invention has  
 CC reduced level of fungal toxins, hence suitable for foodstuffs  
 CC preparation. Yield is high. Plant health is improved due to reduced  
 CC stalk rot. Grain test weight is high and the rate of grain dry down is  
 CC reduced. The present sequence is nopaline synthetase terminator of  
 CC phosphinothricin acetyl transferase (PAT) gene expression cassette.

```
XX SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;
Query Match 100.0%; Score 21; DB 22; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtaacatagatgacaccgcgc 21
Db 247 GTAACATAGATGACACCGCGC 227

RESULT 4
AAD06367/c
ID AAD06367 standard; DNA; 249 BP.
XX
AC AAD06367;
XX
DT 10-AUG-2001 (first entry)
XX
DE 3' nopaline synthase terminator of Pat gene expression cassette.
XX
KW Maize: expression cassette; 35S promoter; alcohol dehydrogenase intron;
KW crystal lAb; CryIAb toxin; phosphinothricin acetyl transferase; PAT;
KW nopaline synthase; NOS terminator; aflatoxin; food material;
KW sweet corn; human food; Adhl-1S; ds.
XX
OS Unidentified.
XX
PN US6229075-B1.
XX
PD 08-MAY-2001.
XX
PF 11-JUN-1999; 99US-0330760.
XX
PR 14-MAR-1997; 97US-0109808.
PR 13-MAR-1998; 98US-0042426.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;
XX
DR WPI; 2001-342708/36.
XX
XX New maize (Zea mays) inbred line R412H (ATCC 209675), useful for
PT producing corn with reduced levels of toxins (e.g. the fungal
PT aflatoxin) and for preparing food materials for human or animal
PT consumption -
XX
XX Claim 3; Column 21-22; 30pp; English.
XX
XX The present invention relates to maize inbred line R412H which comprises
CC a nucleic acid construct with two cassettes, which are transcribed in the
CC same direction. The expression cassette comprises a Cauliflower mosaic
CC virus (CaMV) 35S constitutive promoter operably linked to a maize alcohol
CC dehydrogenase intron, a DNA sequence of a gene encoding a
CC Bacillus thuringiensis (Bt) crystal lAb (CryIAb) toxin or
CC phosphinothricin acetyl transferase (PAT) and a nopaline synthase (NOS)
CC terminator functional in plants. The maize inbred line R412H is useful
CC for producing corn with reduced levels of toxins (e.g. the fungal
CC aflatoxin). This maize line is particularly useful for preparing food
CC materials for human or animal consumption, e.g. sweet corn for
CC packaging or fresh use as human food, or grain or silage made from field
CC corn. The present sequence is 3' NOS terminator of Pat gene expression
CC cassette, which is related to the invention.
XX
XX Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 1 gtaacatagatgacaccgcgc 21
Db 247 GTAACATAGATGACACCGCGC 227

RESULT 5
AAF89824/c
ID AAF89824 standard; DNA; 249 BP.
XX
AC AAF89824;
XX
DT 23-JUL-2001 (first entry)
XX
DE Nucleotide sequence of a 3' terminator from nopaline synthetase.
XX
KW Maize; inbred line R660H; 35S constitutive promoter; CryIAb protein;
KW alcohol dehydrogenase; phosphinothricin acetyl transferase; Chromosome 8;
KW Z1B3; UMC150a; Lepidoptera; gluofosinate; sweet corn; fungal toxin; ds.
XX
OS Unidentified.
XX
PN US6232533-B1.
XX
PD 15-MAY-2001.
XX
PF 09-JUN-1999; 99US-0328473.
XX
PR 14-MAR-1997; 97US-0109808.
PR 13-MAR-1998; 98US-0042426.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;
XX
DR WPI; 2001-335091/35.
XX
XX Novel seed of maize inbred line R372H, useful for producing inbred
PT maize plants which on crossing with other different maize plants
PT produce hybrid maize seeds and plants having resistance to insects and
PT herbicide -
XX
XX Claim 3; Column 19-21; 29pp; English.
XX
XX The specification describes seed and plants of maize inbred line R660H.
CC The seed comprises two cassettes. The first cassette comprises a
CC cauliflower mosaic virus (CaMV) 35S constitutive promoter operably linked
CC to a maize alcohol dehydrogenase intron, a DNA sequence of a gene
CC encoding a CryIAb protein, and a terminator which functions in plants. The
CC second cassette comprises a CaMV 35S promoter which functions in plant
CC cells operably linked to a maize alcohol dehydrogenase intron, a DNA
CC sequence of a gene encoding for phosphinothricin acetyl transferase, and
CC a terminator functional in plant. The two cassettes are transcribed into
CC the same direction and the nucleic acid construct is incorporated into
CC the seed's genome on chromosome 8 and near position 117, between markers
CC Z1B3 and UMC150a. The maize plants and seeds exhibit resistance to
CC Lepidopteran insects since they express the protein CryIAb in high
CC levels, and also exhibit resistance to non-selective herbicide
CC glufosinate. The transgenic maize is suited for preparation of food
CC materials for human or animal consumption e.g. sweet corn for packaging
CC or fresh use as a human food, or grain or silage made from field corn,
CC containing reduced levels of fungal toxins. The present sequence
CC represents a 3' terminator from nopaline synthetase, which is used to
CC produce transgenic maize of the invention.
XX
XX Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtaacatagatgacaccgcgc 21
```





```

PN US6114608-A.
XX
PD 05-SEP-2000.
XX
PF 13-MAR-1998; 98US-0042426.
XX
PR 14-MAR-1997; 97US-0109808.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Dietrich PS, Mettler IJ, Sinibaldi RM;
XX
DR WPI; 2000-586487/55.
XX
PT Novel DNA construct comprising an expression cassette having a
PT functional constitutive promoter, operably linked to a maize alcohol
PT dehydrogenase intron, gene encoding Cry IAb protein and a terminator
PS
PS Claim 1; Column 21-22; 24pp; English.
XX
CC The present sequence represents a nopaline synthase terminator, which is
CC used to create the construct of the invention. The specification
CC describes a nucleic acid construct, comprising an expression cassette
CC containing a cauliflower mosaic virus (CaMV) 35S promoter, a maize
CC alcohol dehydrogenase intron, a DNA molecule encoding an insecticidal
CC Cry IAb protein toxin and a NOS (nopaline synthase) terminator in
CC operable linkage. The nucleic acid constructs are useful for
CC producing transgenic plants such as maize, wheat, sorghum and
CC rice. Preferably, maize plants such as field corn, sweet corn, white
CC corn, silage corn or popcorn are produced. The transgenic maize is
CC used for preparing food materials with reduced levels of fungal toxins,
CC e.g. aflatoxins. The transgenic plants express a baculovirus crystal
CC protein toxin, which is effective against Lepidopteran insects at
CC relatively high levels and further provides resistance to the
CC non-selective herbicide glufosinate.
XX
SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21
DB 253 GTAACATAGATGACACCGCGC 233

RESULT 11
AAD06309/c
ID AAD06309 standard; DNA; 261 BP.
XX
AC AAD06309;
XX
DT 10-AUG-2001 (first entry)
XX
DE Nopaline synthase terminator of Bt kurstaki expression cassette.
XX
KW Transgenic maize; expression cassette; 35S promoter;
KW alcohol dehydrogenase intron; CryIAb toxin; nopaline synthetase;
KW NOS terminator; foodstuffs preparation; ds.
XX
OS Unidentified.
XX
PN US6222104-B1.
XX
PD 24-APR-2001.
XX
PF 13-APR-1999; 99US-0291238.
XX
PR 09-NOV-1994; 94US-0336627.
PR 22-AUG-1996; 96US-0716836.
PR 14-MAR-1997; 97US-0818573.

```

```

PR 13-MAR-1998; 98US-0042426.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Mettler IJ, Krier M, Mies D;
XX
DR WPI; 2001-327266/34.
XX
PT Novel transgenic maize seed for hybrid maize plant production,
PT comprising expression cassette linked operably with CAMV 35S promoter,
PT alcohol dehydrogenase intron, insecticidal CryIAb toxin gene or NOS
PT terminator
XX
PS Claim 1; Column 21-22; 24pp; English.
XX
CC The present invention relates to transgenic maize seed for hybrid maize
CC plant production, comprising an expression cassette linked operably with
CC constitutive Cauliflower mosaic virus (CaMV) 35S promoter, an alcohol
CC dehydrogenase intron, a DNA encoding an insecticidal
CC bacillus thuringiensis (Bt) CryIAb toxin and a nopaline synthetase (NOS)
CC terminator. The transgenic maize seed is used in maize cultivation
CC methods for propagating hybrid maize seeds and for growing hybrid maize
CC plants. The maize is also used in foodstuffs preparation for animal or
CC human consumption. The inbred maize line NP948 of the invention has
CC reduced level of fungal toxins, hence suitable for foodstuffs
CC preparation. Yield is high. Plant health is improved due to reduced
CC stalk rot. Grain test weight is high and the rate of grain dry down is
CC reduced. The present sequence is nopaline synthetase terminator of Bt
CC kurstaki expression cassette.
XX
SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21
DB 253 GTAACATAGATGACACCGCGC 233

RESULT 12
AAD06363/c
ID AAD06363 standard; DNA; 261 BP.
XX
AC AAD06363;
XX
DT 10-AUG-2001 (first entry)
XX
DE 3' nopaline synthase terminator of Btk gene expression cassette.
XX
KW Maize; expression cassette; 35S promoter; alcohol dehydrogenase intron;
KW crystal IAb; CryIAb toxin; phosphinothricin acetyl transferase; PAT;
KW nopaline synthase; NOS terminator; aflatoxin; food material;
KW sweet corn; human food; Adh1-1S; ds.
XX
OS Unidentified.
XX
PN US6229075-B1.
XX
PD 08-MAY-2001.
XX
PF 11-JUN-1999; 99US-0330760.
XX
PR 14-MAR-1997; 97US-0109808.
PR 13-MAR-1998; 98US-0042426.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;
XX
DR WPI; 2001-342708/36.

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:55:52 ; Search time 301.6 Seconds  
(without alignments)  
15.769 Million cell updates/sec

Title: US-09-698-903B-2  
Perfect score: 21  
Sequence: 1 gtaacatagatgacaccgcgc 21

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21	100.0	249	3	US-09-042-426-8
C 2	21	100.0	249	4	US-09-291-238-8
C 3	21	100.0	249	4	US-09-330-760-8
C 4	21	100.0	249	4	US-09-328-473-8
C 5	21	100.0	249	4	US-09-330-737-8
C 6	21	100.0	254	3	US-08-869-696-18
C 7	21	100.0	261	3	US-09-042-426-4
C 8	21	100.0	261	4	US-09-291-238-4
C 9	21	100.0	261	4	US-09-330-760-4
C 10	21	100.0	261	4	US-09-328-473-4
C 11	21	100.0	261	4	US-09-330-737-4
C 12	21	100.0	1287	1	US-08-064-121-3
C 13	21	100.0	1287	1	US-08-478-015-3
C 14	21	100.0	1287	3	US-08-475-975-3
C 15	21	100.0	1287	3	US-09-084-889-3
C 16	21	100.0	1559	3	US-09-049-475-6
C 17	21	100.0	1829	1	US-07-966-187-17
C 18	21	100.0	1863	1	US-08-525-507-16
C 19	21	100.0	1863	2	US-08-475-427-9
C 20	21	100.0	1863	2	US-07-842-165-9
C 21	21	100.0	2115	2	US-08-767-026-3
C 22	21	100.0	2145	1	US-08-592-936B-16
C 23	21	100.0	2145	2	US-09-111-573-16
C 24	21	100.0	2319	3	US-09-049-475-5
C 25	21	100.0	2345	3	US-09-026-673-1
C 26	21	100.0	2633	1	US-08-452-267-2
C 27	21	100.0	2633	3	US-09-123-644-2

C 28	21	100.0	2917	1	US-08-592-936B-20	Sequence 20, Appl
C 29	21	100.0	2917	2	US-09-111-573-20	Sequence 20, Appl
C 30	21	100.0	3183	1	US-08-413-135-1	Sequence 1, Appl
C 31	21	100.0	3237	2	US-08-419-075-26	Sequence 26, Appl
C 32	21	100.0	3544	2	US-08-485-139-3	Sequence 3, Appl
C 33	21	100.0	3544	2	US-08-485-139-3	Sequence 3, Appl
C 34	21	100.0	3544	3	US-08-750-357-3	Sequence 3, Appl
C 35	21	100.0	3544	3	US-08-750-357-3	Sequence 3, Appl
C 36	21	100.0	3658	3	US-08-894-440-3	Sequence 3, Appl
C 37	21	100.0	3658	3	US-08-894-440-3	Sequence 3, Appl
C 38	21	100.0	4284	1	US-08-525-507-14	Sequence 14, Appl
C 39	21	100.0	4583	3	US-08-810-720-9	Sequence 9, Appl
C 40	21	100.0	4808	1	US-08-351-413-17	Sequence 17, Appl
C 41	21	100.0	4808	1	US-08-351-413-17	Sequence 17, Appl
C 42	21	100.0	4808	2	US-09-025-583-17	Sequence 17, Appl
C 43	21	100.0	4808	2	US-09-025-583-17	Sequence 17, Appl
C 44	21	100.0	4883	1	US-08-064-121-4	Sequence 4, Appl
C 45	21	100.0	4883	3	US-09-084-889-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-09-042-426-8/c  
; Sequence 8, Application US/09042426  
; Patent No. 6114608  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6114608artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS Terminator  
; US-09-042-426-8

Query Match 100.0%; Score 21; DB 3; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 gtaacatagatgacaccgcgc 21

Db 247 GTAACATAGATGACACCGGC 227  
|||||

## RESULT 2

US-09-291-238-8/c  
; Sequence 8, Application US/09291238  
; Patent No. 6222104  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6222104artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/291,238  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/042,426  
; FILING DATE: March 13, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS Terminator  
US-09-291-238-8

Query Match 100.0%; Score 21; DB 4; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGGC 227

## RESULT 3

US-09-330-760-8/c  
; Sequence 8, Application US/09330760  
; Patent No. 6229075  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6229075artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit

; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/330,760  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/042,426  
; FILING DATE: March 13, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS Terminator  
US-09-330-760-8

Query Match 100.0%; Score 21; DB 4; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGGC 227

## RESULT 4

US-09-328-473-8/c  
; Sequence 8, Application US/09328473  
; Patent No. 6232533  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6232533artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/328,473  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/042,426  
; FILING DATE: March 13, 1998  
; ATTORNEY/AGENT INFORMATION:

; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS Terminator  
; US-09-328-473-8

Query Match 100.0%; Score 21; DB 4; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||  
DB 247 GTAACATAGATGACACCGCGC 227

RESULT 5  
US-09-330-737-8/c  
; Sequence 8, Application US/09330737  
; Patent No. 6232534  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6232534artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/09/330,737  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:

; CLONE: NOS Terminator  
; US-09-330-737-8

Query Match 100.0%; Score 21; DB 4; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||  
DB 247 GTAACATAGATGACACCGCGC 227

RESULT 6  
US-08-869-696-18/c  
; Sequence 18, Application US/08869696C  
; Patent No. 6031155  
; GENERAL INFORMATION:  
; APPLICANT: Cameron-Mills, Verena  
; APPLICANT: Lok, Finn  
; APPLICANT: Sinjorgo, Catharina Maria Cornelia  
; APPLICANT: Van Den Dool, Ronald Tako Marinus  
; APPLICANT: Caspers, Martinus Petrus Maria  
; APPLICANT: Van Zuijl-Van Der Valk, Maria Joanna  
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION  
; FILE REFERENCE: 11225.01US01  
; CURRENT APPLICATION NUMBER: US/08/869,696C  
; CURRENT FILING DATE: 1997-06-05  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 254  
; TYPE: DNA  
; ORGANISM: synthetic  
; US-08-869-696-18

Query Match 100.0%; Score 21; DB 3; Length 254;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||  
DB 247 GTAACATAGATGACACCGCGC 227

RESULT 7  
US-09-042-426-4/c  
; Sequence 4, Application US/09042426  
; Patent No. 6114608  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6114608artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993

REFERENCE/DOCKET NUMBER: 135/1  
TELEPHONE: (919) 541-8614  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: NOS terminator  
US-09-042-426-4

Query Match 100.0%; Score 21; DB 3; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 8  
US-09-291-238-4/c  
Sequence 4, Application US/09291238  
Patent No. 622104  
GENERAL INFORMATION:  
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
TITLE OF INVENTION: DNA Construct Containing Bacillus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 622104artis Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/291,238  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/042,426  
FILING DATE: March 13, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoxie, Thomas  
REGISTRATION NUMBER: 32,993  
REFERENCE/DOCKET NUMBER: 135/1  
TELEPHONE: (919) 541-8614  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: NOS terminator  
US-09-291-238-4

Query Match 100.0%; Score 21; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 9  
US-09-330-760-4/c  
Sequence 4, Application US/09330760  
Patent No. 6229075  
GENERAL INFORMATION:  
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
TITLE OF INVENTION: DNA Construct Containing Bacillus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6229075artis Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/330,760  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/042,426  
FILING DATE: March 13, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoxie, Thomas  
REGISTRATION NUMBER: 32,993  
REFERENCE/DOCKET NUMBER: 135/1  
TELEPHONE: (919) 541-8614  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: NOS terminator  
US-09-330-760-4

Query Match 100.0%; Score 21; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 10  
US-09-328-473-4/c  
Sequence 4, Application US/09328473  
Patent No. 6232533  
GENERAL INFORMATION:

```

; APPLICATION NUMBER: US/09/330,737
; FILING DATE:
; CLASSIFICATION:
; PRIORITY INFORMATION DATA:
; PRIOR APPLICATION NUMBER: US/09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoxie, Thomas
; REGISTRATION NUMBER: 32,993
; REFERENCE/DOCKET NUMBER: 135/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8614
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: NOS terminator
;
US-09-330-737-4

Query Match 100.0%; Score 21
Best Local Similarity 100.0%; Pred. NO
Matches 21; Conservative 0; Mismatch

Qy 1 gtaacatagatgacacgcgc 21
|||||
Db 253 GTAACATAGTACACGC GC 233

RESULT 12
US-08-064-121-3/c
; Sequence 3, Application US/08064121
; Patent No. 5641664
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANS-
; TITLE OF INVENTION: MONOCOTYLEDONOUS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker &
; STREET: George Mason Bldg., Washing
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Ve
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/064,121
; FILING DATE: 24-MAY-1993
; CLASSIFICATION: 800
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:

```

APPLICANT: GOBEL, ELKE  
 TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
 TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: George Mason Bldg., Washington & Prince Sts.  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/064,121  
 FILING DATE: 24-MAY-1993  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: EP 90403332.1  
 FILING DATE: 23-NOV-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: EP 91401888.2  
 FILING DATE: 08-JUL-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Crane-Feury, Sharon E  
 REGISTRATION NUMBER: 36,113  
 REFERENCE/DOCKET NUMBER: 010830-043  
 TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= pTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
; US-08-064-121-3

Query Match 100.0%; Score 21; DB 1; Length 1287;
Best Local Similarity 100.0%; Pred. NO. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 1251 GTAACATAGATGACACGC 1231

```

```

RESULT 13
US-08-478-015-3/c
; Sequence 3, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= pTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylat
; OTHER INFORMATION: site derived from Agrobacterium T-DNA nopaline synthase ge
; US-08-478-015-3

Query Match 100.0%; Score 21; DB 1; Length 1287;
Best Local Similarity 100.0%; Pred. NO. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 1251 GTAACATAGATGACACGC 1231

RESULT 14
US-08-475-975-3/c
; Sequence 3, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25

```



```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475.975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= pTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
;
US-08-475-975-3

Query Match 100.0%; Score 21; DB 3; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
   ||||||||||||||||||
Db 1251 GTACATAGATGACACCGCGC 1231

RESULT 15
US-09-084-889-3/c
; Sequence 3, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.

```

```

;
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/064,121
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= pTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
;
US-09-084-889-3

Query Match 100.0%; Score 21; DB 3; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
   ||||||||||||||||||
Db 1251 GTACATAGATGACACCGCGC 1231

Search completed: February 25, 2002, 18:05:08
Job time: 18556 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:51:47 ; Search time 8261.74 Seconds  
(without alignments)  
27.314 Million cell updates/sec

Title: US-09-698-903B-2

Perfect score: 21

Sequence: 1 gtaacatagatgacaccgcgc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estfun:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estom:\*
  - 5: em\_estpl:\*
  - 6: em\_estba:\*
  - 7: em\_estro:\*
  - 8: em\_estov:\*
  - 9: em\_htc:\*
  - 10: gb\_est1:\*
  - 11: gb\_est2:\*
  - 12: gb\_htc:\*
  - 13: gb\_gss:\*
  - 14: em\_gss\_fun:\*
  - 15: em\_gss\_hum:\*
  - 16: em\_gss\_inv:\*
  - 17: em\_gss\_pln:\*
  - 18: em\_gss\_pro:\*
  - 19: em\_gss\_rod:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	17.8	84.8	887	13	AQ363458
c 2	17.4	82.9	300	10	AU114795
c 3	17.4	82.9	833	11	BF104658
c 4	17.4	82.9	936	13	CNS01EVA
5	17	81.0	215	10	BB143973
6	17	81.0	277	10	BB244247
7	16.8	80.0	232	10	AV008164
8	16.8	80.0	379	10	AV684098
9	16.8	80.0	542	13	CNS00WTL
10	16.8	80.0	594	10	AI109780
11	16.8	80.0	662	13	AZ029768
12	16.8	80.0	710	10	BE038757

13	16.8	80.0	826	10	AW940028
14	16.8	80.0	910	11	BG823293
c 15	16.4	78.1	245	10	AA359059
c 16	16.4	78.1	282	10	BB383808
c 17	16.4	78.1	291	10	AA361337
c 18	16.4	78.1	299	10	AI081942
c 19	16.4	78.1	343	10	BE118822
c 20	16.4	78.1	348	11	BE1013588
c 21	16.4	78.1	411	11	N71737
c 22	16.4	78.1	423	11	R58585
c 23	16.4	78.1	435	10	BE710208
c 24	16.4	78.1	442	10	AV617102
c 25	16.4	78.1	442	10	AV617103
c 26	16.4	78.1	455	11	W92829
c 27	16.4	78.1	460	10	AW652485
c 28	16.4	78.1	470	10	AV725000
c 29	16.4	78.1	478	10	AL121482
c 30	16.4	78.1	494	10	AA010698
c 31	16.4	78.1	577	11	BF342064
c 32	16.4	78.1	612	11	BF038038
c 33	16.4	78.1	632	10	AW607872
c 34	16.4	78.1	643	11	BG705748
c 35	16.4	78.1	657	10	AW853029
c 36	16.4	78.1	692	11	BG702664
c 37	16.4	78.1	693	11	BG422864
c 38	16.4	78.1	699	11	BG714597
c 39	16.4	78.1	704	11	BG747014
c 40	16.4	78.1	742	10	BE409302
c 41	16.4	78.1	747	11	BG325494
c 42	16.4	78.1	750	11	BG701119
c 43	16.4	78.1	802	10	AL546504
c 44	16.4	78.1	834	11	BG716798
c 45	16.4	78.1	851	10	AU117875

ALIGNMENTS

RESULT 1  
LOCUS AQ363458/887 bp DNA GSS 16-DEC-1999  
DEFINITION nbxb0059A08f CUGI Rice BAC Library Oryza sativa genomic clone  
ACCESSION nbxb0059A08f, DNA sequence.  
VERSION AQ363458  
KEYWORDS GSS.  
SOURCE AQ363458.2 GI:6583332  
ORGANISM Oryza sativa.  
REFERENCE 1 (bases 1 to 887)  
AUTHORS Wing, R.A. and Dean, R.A.  
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
JOURNAL Unpublished (1998)  
COMMENT On Dec 15, 1999 this sequence version replaced gi:4213113.  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCCTATAGGG  
Class: BAC ends  
High quality sequence stop: 1.  
Location/Qualifiers  
1. .887  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbxb0059A08f"

/clone\_lib="CUGI Rice BAC Library"

/tissue\_type="Leaf"

/lab\_host="E. coli DH10B"

/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 260 a 200 c 189 g 236 t 2 others

ORIGIN

Query Match 84.8%; Score 17.8; DB 13; Length 887;  
Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21

||||| ||||| ||||| |||||

Db 517 GCAACATTGATGACACGCGC 497

RESULT 2

AU114795/c

LOCUS

DEFINITION AU114795 300 bp mRNA EST 19-OCT-2000  
elegans cDNA clone yk728e10 3', mRNA sequence.

ACCESSION AU114795

VERSION AU114795.1

KEYWORDS GI:10928362

SOURCE EST.

ORGANISM Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 300)

REFERENCE Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.

and Sugano, S.

A complementary view of the C. elegans genome

Unpublished (2000)

Contact: Yuji Kohara

Genome Biology Lab.

National Institute of Genetics

Yata Iiri, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

Location/Qualifiers

1..300

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk728e10"

/clone\_lib="unpublished oligo-capped cDNA library"

/sex="Hermaphrodite"

/tissue\_type="whole animal"

/dev\_stage="varied"

96 a 49 c 66 g 86 t 3 others

BASE COUNT

ORIGIN

Query Match 82.9%; Score 17.4; DB 10; Length 300;  
Best Local Similarity 94.7%; Pred. No. 1.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 19

||||| ||||| ||||| |||||

Db 148 GTAACATAGATGACACGCGC 130

RESULT 3

BF104658/c

LOCUS

DEFINITION BF104658 833 bp mRNA EST 19-OCT-2000

601647679R1 NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:4041811 3',

mRNA sequence.

ACCESSION BF104658

VERSION BF104658.1

KEYWORDS GI:10887184

SOURCE EST.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 833)

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: CLONETECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM861 row: n column: 20.

Location/Qualifiers

1..833

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4041811"

/clone\_lib="NIH\_MGC\_61"

/tissue\_type="embryonal carcinoma"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: Testis; Vector: pDNR-LIB (Clontech); Site\_1:

SfiI (ggccattggcgc); Site\_2: SfiI (ggccattggcgc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3'

sequence: 5'-CACGGCATTGGCC-3' and 3' adaptor

(where B = A, C, G, or T).

Average

insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC

Library."

BASE COUNT 199 a 213 c 196 g 225 t

ORIGIN

Query Match 82.9%; Score 17.4; DB 11; Length 833;  
Best Local Similarity 94.7%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taacatagatgacacgcgc 20

||||| ||||| ||||| |||||

Db 600 TAACATAGATGACGCGCGC 582

RESULT 4

CNS01EY4

LOCUS

DEFINITION CNS01EY4 936 bp DNA GSS 30-MAY-2001

Anopheles gambiae GSS T7 end of clone 03E09 of Notredamei library

from strain PEST of Anopheles gambiae (African malaria mosquito),



## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 277)

## REFERENCE

Authors  
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci  
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
Izawa, M., Kadota, K., Kadowa, I., Kai, C., Kawai, J., Kikuchi, N.,  
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata  
Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suganara, Y.,  
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya  
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,  
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino  
M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)

## TITLE

RIKEN Mouse ESTs (Konno, H., et al.)

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Tomaru, Y., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki  
Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## FEATURES

## source

Location/Qualifiers  
1. .277  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A73001C13"  
/clone\_lib="RIKEN full-length enriched, 7 days neonate  
cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="7 days neonate"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI: cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGACGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 20.0 and subtraction to Rot = 459.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGACGAGATCTCGAGTAAATTAATTCCTCCCTCCCTCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(-) after bulk excision from Lambda  
FLC I."

## BASE COUNT

ORIGIN  
83 a 49 c 55 g 90 t

## Query Match

Best Local Similarity 81.0%; Score 17; DB 10; Length 277;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 taacatagatgacaccg 18

DB 65 TAACATAGATGACACCG 81

## RESULT

AV008164

LOCUS

DEFINITION

AV008164

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

## TITLE

JOURNAL

COMMENT

## FEATURES

## source

Location/Qualifiers  
1. .232  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1110008H08"  
/clone\_lib="Mus musculus 18-day embryo C57BL/6J"  
/sex="mixed"  
/dev\_stage="18-day embryo"  
BASE COUNT 84 a 32 c 45 g 71 t  
ORIGIN

## Query Match

Best Local Similarity 80.0%; Score 16.8; DB 10; Length 232;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccg 20

DB 102 GTAACACAGATGACACGCG 121

## RESULT

AV684098

LOCUS

DEFINITION

ACCESSION

AV684098 379 bp mRNA EST 25-SEP-2000

AV684098 GKC Homo sapiens cDNA clone GKCBQ09 5', mRNA sequence.

AV684098



```

RESULT 11
AZ029768
LOCUS
DEFINITION
  AZ029768 662 bp DNA GSS 25-FEB-2000
  RPCI-23-349G17.TV RPCI-23 Mus musculus genomic clone RPCI-23-349G17
  DNA sequence.
ACCESSION
  AZ029768
VERSION
  1
KEYWORDS
  GSS.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
REFERENCE
  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 662)
AUTHORS
  Zhao,S., Nierman,W., Feidblyum,T., Malek,J., Shatsman,S., Akinret
  ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
  and Fraser,C.M.
TITLE
  Mouse BAC End Sequences from Library RPCI-23
JOURNAL
  Unpublished (1999)
COMMENT
  Other_GSSs: RPCI-23-349G17.TJ
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-23. For BAC
  library availability, please contact Pieter de Jong
  (pieter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
  or from Resea ch Genetics (info@resgen.com). BAC end page:
  http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
  Plate: 349 row: G column: 17
  Seq primer: T7
  Class: BAC ends.
FEATURES
  source
    Location/Qualifiers
      1..662
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="RPCI-23-349G17"
        /clone_lib="RPCI-23"
        /sex="Female"
        /lab_host="DH10B"
        /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
        EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
        brain genomic DNA was isolated and partially digested
        with a combination of EcoRI and EcoRI. Size
        selected DNA was cloned into the pBACe3.6 vector at the
        EcoRI sites. The ligation products were transformed into
        DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 180 a 136 c 175 g 170 t
ORIGIN
  1
  100
  200
  300
  400
  500
  600
  700
  800
  900
  1000
  1100
  1200
  1300
  1400
  1500
  1600
  1700
  1800
  1900
  2000
  2100
  2200
  2300
  2400
  2500
  2600
  2700
  2800
  2900
  3000
  3100
  3200
  3300
  3400
  3500
  3600
  3700
  3800
  3900
  4000
  4100
  4200
  4300
  4400
  4500
  4600
  4700
  4800
  4900
  5000
  5100
  5200
  5300
  5400
  5500
  5600
  5700
  5800
  5900
  6000
  6100
  6200
  6300
  6400
  6500
  6600
  6700
  6800
  6900
  7000
  7100
  7200
  7300
  7400
  7500
  7600
  7700
  7800
  7900
  8000
  8100
  8200
  8300
  8400
  8500
  8600
  8700
  8800
  8900
  9000
  9100
  9200
  9300
  9400
  9500
  9600
  9700
  9800
  9900
  10000
  10100
  10200
  10300
  10400
  10500
  10600
  10700
  10800
  10900
  11000
  11100
  11200
  11300
  11400
  11500
  11600
  11700
  11800
  11900
  12000
  12100
  12200
  12300
  12400
  12500
  12600
  12700
  12800
  12900
  13000
  13100
  13200
  13300
  13400
  13500
  13600
  13700
  13800
  13900
  14000
  14100
  14200
  14300
  14400
  14500
  14600
  14700
  14800
  14900
  15000
  15100
  15200
  15300
  15400
  15500
  15600
  15700
  15800
  15900
  16000
  16100
  16200
  16300
  16400
  16500
  16600
  16700
  16800
  16900
  17000
  17100
  17200
  17300
  17400
  17500
  17600
  17700
  17800
  17900
  18000
  18100
  18200
  18300
  18400
  18500
  18600
  18700
  18800
  18900
  19000
  19100
  19200
  19300
  19400
  19500
  19600
  19700
  19800
  19900
  20000
  20100
  20200
  20300
  20400
  20500
  20600
  20700
  20800
  20900
  21000
  21100
  21200
  21300
  21400
  21500
  21600
  21700
  21800
  21900
  22000
  22100
  22200
  22300
  22400
  22500
  22600
  22700
  22800
  22900
  23000
  23100
  23200
  23300
  23400
  23500
  23600
  23700
  23800
  23900
  24000
  24100
  24200
  24300
  24400
  24500
  24600
  24700
  24800
  24900
  25000
  25100
  25200
  25300
  25400
  25500
  25600
  25700
  25800
  25900
  26000
  26100
  26200
  26300
  26400
  26500
  26600
  26700
  26800
  26900
  27000
  27100
  27200
  27300
  27400
  27500
  27600
  27700
  27800
  27900
  28000
  28100
  28200
  28300
  28400
  28500
  28600
  28700
  28800
  28900
  29000
  29100
  29200
  29300
  29400
  29500
  29600
  29700
  29800
  29900
  30000
  30100
  30200
  30300
  30400
  30500
  30600
  30700
  30800
  30900
  31000
  31100
  31200
  31300
  31400
  31500
  31600
  31700
  31800
  31900
  32000
  32100
  32200
  32300
  32400
  32500
  32600
  32700
  32800
  32900
  33000
  33100
  33200
  33300
  33400
  33500
  33600
  33700
  33800
  33900
  34000
  34100
  34200
  34300
  34400
  34500
  34600
  34700
  34800
  34900
  35000
  35100
  35200
  35300
  35400
  35500
  35600
  35700
  35800
  35900
  36000
  36100
  36200
  36300
  36400
  36500
  36600
  36700
  36800
  36900
  37000
  37100
  37200
  37300
  37400
  37500
  37600
  37700
  37800
  37900
  38000
  38100
  38200
  38300
  38400
  38500
  38600
  38700
  38800
  38900
  39000
  39100
  39200
  39300
  39400
  39500
  39600
  39700
  39800
  39900
  40000
  40100
  40200
  40300
  40400
  40500
  40600
  40700
  40800
  40900
  41000
  41100
  41200
  41300
  41400
  41500
  41600
  41700
  41800
  41900
  42000
  42100
  42200
  42300
  42400
  42500
  42600
  42700
  42800
  42900
  43000
  43100
  43200
  43300
  43400
  43500
  43600
  43700
  43800
  43900
  44000
  44100
  44200
  44300
  44400
  44500
  44600
  44700
  44800
  44900
  45000
  45100
  45200
  45300
  45400
  45500
  45600
  45700
  45800
  45900
  46000
  46100
  46200
  46300
  46400
  46500
  46600
  46700
  46800
  46900
  47000
  47100
  47200
  47300
  47400
  47500
  47600
  47700
  47800
  47900
  48000
  48100
  48200
  48300
  48400
  48500
  48600
  48700
  48800
  48900
  49000
  49100
  49200
  49300
  49400
  49500
  49600
  49700
  49800
  49900
  50000
  50100
  50200
  50300
  50400
  50500
  50600
  50700
  50800
  50900
  51000
  51100
  51200
  51300
  51400
  51500
  51600
  51700
  51800
  51900
  52000
  52100
  52200
  52300
  52400
  52500
  52600
  52700
  52800
  52900
  53000
  53100
  53200
  53300
  53400
  53500
  53600
  53700
  53800
  53900
  54000
  54100
  54200
  54300
  54400
  54500
  54600
  54700
  54800
  54900
  55000
  55100
  55200
  55300
  55400
  55500
  55600
  55700
  55800
  55900
  56000
  56100
  56200
  56300
  56400
  56500
  56600
  56700
  56800
  56900
  57000
  57100
  57200
  57300
  57400
  57500
  57600
  57700
  57800
  57900
  58000
  58100
  58200
  58300
  58400
  58500
  58600
  58700
  58800
  58900
  59000
  59100
  59200
  59300
  59400
  59500
  59600
  59700
  59800
  59900
  60000
  60100
  60200
  60300
  60400
  60500
  60600
  60700
  60800
  60900
  61000
  61100
  61200
  61300
  61400
  61500
  61600
  61700
  61800
  61900
  62000
  62100
  62200
  62300
  62400
  62500
  62600
  62700
  62800
  62900
  63000
  63100
  63200
  63300
  63400
  63500
  63600
  63700
  63800
  63900
  64000
  64100
  64200
  64300
  64400
  64500
  64600
  64700
  64800
  64900
  65000
  65100
  65200
  65300
  65400
  65500
  65600
  65700
  65800
  65900
  66000
  66100
  66200
  66300
  66400
  66500
  66600
  66700
  66800
  66900
  67000
  67100
  67200
  67300
  67400
  67500
  67600
  67700
  67800
  67900
  68000
  68100
  68200
  68300
  68400
  68500
  68600
  68700
  68800
  68900
  69000
  69100
  69200
  69300
  69400
  69500
  69600
  69700
  69800
  69900
  70000
  70100
  70200
  70300
  70400
  70500
  70600
  70700
  70800
  70900
  71000
  71100
  71200
  71300
  71400
  71500
  71600
  71700
  71800
  71900
  72000
  72100
  72200
  72300
  72400
  72500
  72600
  72700
  72800
  72900
  73000
  73100
  73200
  73300
  73400
  73500
  73600
  73700
  73800
  73900
  74000
  74100
  74200
  74300
  74400
  74500
  74600
  74700
  74800
  74900
  75000
  75100
  75200
  75300
  75400
  75500
  75600
  75700
  75800
  75900
  76000
  76100
  76200
  76300
  76400
  76500
  76600
  76700
  76800
  76900
  77000
  77100
  77200
  77300
  77400
  77500
  77600
  77700
  77800
  77900
  78000
  78100
  78200
  78300
  78400
  78500
  78600
  78700
  78800
  78900
  79000
  79100
  79200
  79300
  79400
  79500
  79600
  79700
  79800
  79900
  80000
  80100
  80200
  80300
  80400
  80500
  80600
  80700
  80800
  80900
  81000
  81100
  81200
  81300
  81400
  81500
  81600
  81700
  81800
  81900
  82000
  82100
  82200
  82300
  82400
  82500
  82600
  82700
  82800
  82900
  83000
  83100
  83200
  83300
  83400
  83500
  83600
  83700
  83800
  83900
  84000
  84100
  84200
  84300
  84400
  84500
  84600
  84700
  84800
  84900
  85000
  85100
  85200
  85300
  85400
  85500
  85600
  85700
  85800
  85900
  86000
  86100
  86200
  86300
  86400
  86500
  86600
  86700
  86800
  86900
  87000
  87100
  87200
  87300
  87400
  87500
  87600
  87700
  87800
  87900
  88000
  88100
  88200
  88300
  88400
  88500
  88600
  88700
  88800
  88900
  89000
  89100
  89200
  89300
  89400
  89500
  89600
  89700
  89800
  89900
  90000
  90100
  90200
  90300
  90400
  90500
  90600
  90700
  90800
  90900
  91000
  91100
  91200
  91300
  91400
  91500
  91600
  91700
  91800
  91900
  92000
  92100
  92200
  92300
  92400
  92500
  92600
  92700
  92800
  92900
  93000
  93100
  93200
  93300
  93400
  93500
  93600
  93700
  93800
  93900
  94000
  94100
  94200
  94300
  94400
  94500
  94600
  94700
  94800
  94900
  95000
  95100
  95200
  95300
  95400
  95500
  95600
  95700
  95800
  95900
  96000
  96100
  96200
  96300
  96400
  96500
  96600
  96700
  96800
  96900
  97000
  97100
  97200
  97300
  97400
  97500
  97600
  97700
  97800
  97900
  98000
  98100
  98200
  98300
  98400
  98500
  98600
  98700
  98800
  98900
  99000
  99100
  99200
  99300
  99400
  99500
  99600
  99700
  99800
  99900
  100000
  100100
  100200
  100300
  100400
  100500
  100600
  100700
  100800
  100900
  101000
  101100
  101200
  101300
  101400
  101500
  101600
  101700
  101800
  101900
  102000
  102100
  102200
  102300
  102400
  102500
  102600
  102700
  102800
  102900
  103000
  103100
  103200
  103300
  103400
  103500
  103600
  103700
  103800
  103900
  104000
  104100
  104200
  104300
  104400
  104500
  104600
  104700
  104800
  104900
  105000
  105100
  105200
  105300
  105400
  105500
  105600
  105700
  105800
  105900
  106000
  106100
  106200
  106300
  106400
  106500
  106600
  106700
  106800
  106900
  107000
  107100
  107200
  107300
  107400
  107500
  107600
  107700
  107800
  107900
  108000
  108100
  108200
  108300
  108400
  108500
  108600
  108700
  108800
  108900
  109000
  109100
  109200
  109300
  109400
  109500
  109600
  109700
  109800
  109900
  110000
  110100
  110200
  110300
  110400
  110500
  110600
  110700
  110800
  110900
  111000
  111100
  111200
  111300
  111400
  111500
  111600
  111700
  111800
  111900
  112000
  112100
  112200
  112300
  112400
  112500
  112600
  112700
  112800
  112900
  113000
  113100
  113200
  113300
  113400
  113500
  113600
  113700
  113800
  113900
  114000
  114100
  114200
  114300
  114400
  114500
  114600
  114700
  114800
  114900
  115000
  115100
  115200
  115300
  115400
  115500
  115600
  115700
  115800
  115900
  116000
  116100
  116200
  116300
  116400
  116500
  116600
  116700
  116800
  116900
  117000
  117100
  117200
  117300
  117400
  117500
  117600
  117700
  117800
  117900
  118000
  118100
  118200
  118300
  118400
  118500
  118600
  118700
  118800
  118900
  119000
  119100
  119200
  119300
  119400
  119500
  119600
  119700
  119800
  119900
  120000
  120100
  120200
  120300
  120400
  120500
  120600
  120700
  120800
  120900
  121000
  121100
  121200
  121300
  121400
  121500
  121600
  121700
  121800
  121900
  122000
  122100
  122200
  122300
  122400
  122500
  122600
  122700
  122800
  122900
  123000
  123100
  123200
  123300
  123400
  123500
  123600
  123700
  123800
  123900
  124000
  124100
  124200
  124300
  124400
  124500
  124600
  124700
  124800
  124900
  125000
  125100
  125200
  125300
  125400
  125500
  125600
  125700
  125800
  125900
  126000
  126100
  126200
  126300
  126400
  126500
  126600
  126700
  126800
  126900
  127000
  127100
  127200
  127300
  127400
  127500
  127600
  127700
  127800
  127900
  128000
  128100
  128200
  128300
  128400
  128500
  128600
  128700
  128800
  128900
  129000
  129100
  129200
  129300
  129400
  129500
  129600
  129700
  129800
  129900
  130000
  130100
  130200
  130300
  130400
  130500
  130600
  130700
  130800
  130900
  131000
  131100
  131200
  131300
  131400
  131500
  131600
  131700
  131800
  131900
  132000
  132100
  132200
  132300
  132400
  132500
  132600
  132700
  132800
  132900
  133000
  133100
  133200
  133300
  133400
  133500
  133600
  133700
  133800
  133900
  134000
  134100
  134200
  134300
  134400
  134500
  134600
  134700
  134800
  134900
  135000
  135100
  135200
  135300
  135400
  135500
  135600
  135700
  135800
  135900
  136000
  136100
  136200
  136300
  136400
  136500
  136600
  136700
  136800
  136900
  137000
  137100
  137200
  137300
  137400
  137500
  137600
  137700
  137800
  137900
  138000
  138100
  138200
  138300
  138400
  138500
  138600
  138700
  138800
  138900
  139000
  1391
```





Search completed: February 25, 2002, 17:20:51  
Job time: 16144 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:59:54 ; Search time 2331.3 Seconds  
(without alignments)  
148.604 Million cell updates/sec

Title: US-09-698-903B-3

Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_on.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	100.0	21	6	AX127750	AX127750 Sequence
2	21	100.0	3266	6	A08065	A08065 Synthetic n
3	21	100.0	3266	6	A10061	A10061 N.tabacum (
c 4	21	100.0	4832	6	AX172441	AX172441 Sequence
c 5	21	100.0	4946	6	A60108	A60108 Sequence 1
c 6	21	100.0	4946	6	A76915	A76915 Sequence 1
c 7	21	100.0	4946	6	AR098307	AR098307 Sequence
c 8	21	100.0	4946	6	AX172440	AX172440 Sequence
c 9	21	100.0	5865	6	AX127748	AX127748 Sequence
c 10	21	100.0	6254	8	NTA29	X52283 Tobacco ant
c 11	19.4	92.4	3243	5	XL070985	U70985 Xenopus lae
c 12	19.4	92.4	149710	2	AC022563	AC022563 Homo sapi
c 13	19.4	92.4	179183	2	AC019254	AC019254 Homo sapi
c 14	17.8	84.8	167409	2	AC073374	AC073374 Homo sapi
c 15	17.8	84.8	178510	2	AC013820	AC013820 Homo sapi
c 16	17.4	82.9	39443	9	AC010505	AC010505 Homo sapi
c 17	17.4	82.9	172246	2	AC024941	AC024941 Homo sapi
c 18	17.4	82.9	195516	2	AC053540	AC053540 Homo sapi
c 19	17.4	82.9	207856	2	AC078778	AC078778 Homo sapi
c 20	17	81.0	4900	5	XLAPICAL	Z14997 X.laavis mr
c 21	17	81.0	123016	2	AF189001	AF189001 Homo sapi
c 22	17	81.0	159934	9	AL160273	AL160273 Human DNA
c 23	17	81.0	182942	2	AC087793	AC087793 Homo sapi
c 24	17	81.0	186413	2	AL161913	AL161913 Homo sapi
c 25	16.8	80.0	1485	14	AF195032	AF195032 Dengue vi
c 26	16.8	80.0	1485	14	AF195033	AF195033 Dengue vi
c 27	16.8	80.0	1485	14	AF195034	AF195034 Dengue vi
c 28	16.8	80.0	1485	14	AF195035	AF195035 Dengue vi
c 29	16.8	80.0	1485	14	AF195036	AF195036 Dengue vi
c 30	16.8	80.0	1485	14	AF195037	AF195037 Dengue vi
c 31	16.8	80.0	1485	14	AF195038	AF195038 Dengue vi
c 32	16.8	80.0	1485	14	AF195039	AF195039 Dengue vi
c 33	16.8	80.0	1485	14	AF264053	AF264053 Dengue vi
c 34	16.8	80.0	1485	14	DENENVB	M24445 Dengue viru
c 35	16.8	80.0	1485	14	DENENVC	M24446 Dengue viru
c 36	16.8	80.0	1485	14	DENENVV	M24447 Dengue viru
c 37	16.8	80.0	1485	14	DVU34933	U34933 Dengue viru
c 38	16.8	80.0	1485	14	DVU34934	U34934 Dengue viru
c 39	16.8	80.0	1485	14	DVU34935	U34935 Dengue viru
c 40	16.8	80.0	1485	14	DVU34936	U34936 Dengue viru
c 41	16.8	80.0	1485	14	DVU34937	U34937 Dengue viru
c 42	16.8	80.0	1485	14	DVU34938	U34938 Dengue viru
c 43	16.8	80.0	1485	14	DVU34939	U34939 Dengue viru
c 44	16.8	80.0	1485	14	DVU34940	U34940 Dengue viru
c 45	16.8	80.0	1485	14	TOGDENM2	X15433 Dengue viru

ALIGNMENTS

RESULT 1	AX127750	AX127750	21 bp	DNA	PAT	15-MAY-2001
LOCUS	Sequence 3	from Patent WO0131042.				
DEFINITION	AX127750					
ACCESSION	AX127750.1	GI:14134397				
VERSION						
KEYWORDS	synthetic construct.					
SOURCE	synthetic construct					
ORGANISM	artificial sequence.					
REFERENCE	1 (bases 1 to 21)					
AUTHORS	Weston,B. and de Beuckeleer,M.					
TITLE	Male-sterile brassica plants and methods for producing same					
JOURNAL	Patent: WO 0131042-A 3 03-MAY-2001;					
FEATURES	Aventis CropScience N.V. (BE)					
source	Location/Qualifiers					
	1..21					
	/organism="synthetic construct"					
	/db_xref="taxon:32630"					
	/note="primer MLD008"					

```
BASE COUNT      4 a      1 c      10 g      6 t
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 1 ATAGGTGGGAGGCTATTGG 21

RESULT 2
A08065      A08065      3266 bp      DNA      PAT      12-AUG-1993
LOCUS      Synthetic nucleotide sequence of the TA29 gene.
DEFINITION
ACCESSION      A08065
VERSION      A08065.1 GI:413313
KEYWORDS
SOURCE      synthetic construct.
ORGANISM      synthetic construct.
REFERENCE      1 (bases 1 to 3266)
AUTHORS
JOURNAL
FEATURES
Source      Location/Qualifiers
1..3266
/organism="synthetic construct"
/db_xref="taxon:32630"
1527..2492
/gene="TA29"
1527..2492
/gene="TA29"
/codon_start=1
/protein_id="CAA00738.1"
/db_xref="GI:413314"
/translation="MVAPKWFISPMILLSLAICSGQPVTSDAIKAKADHNLKAHT
LSNIDAKFGGGGGGIGGAGGAGGGGGGSDTPNVGNPGCSIHGCTVPGGFL
PKPVFVPVSPGCGYCPADIPGTGGMPTESKITGISQARLYRCKPNNMCDKDCNE
LLLFVFMQDKHDKQHLRYGRRGIGLTVGGVGGFIFGFGAMGGGGGGGGSDA
PGCSNDGDPGCGPGGYACPAANNPGSGGITEFHISGLSRFDGYPYRCRPMCESEDC
NELLLHFVSPMQHKHNRHDIHVERSDDEAAHQSKQKHDEDIIN"

BASE COUNT      1016 a      581 c      623 g      1046 t
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 3266;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 178 ATAGGTGGGAGGCTATTGG 198

RESULT 4
AX172441/c      AX172441      4832 bp      DNA      PAT      03-JUL-2001
LOCUS      Sequence 2 from Patent WO0141558.
DEFINITION      AX172441
ACCESSION      AX172441
VERSION      AX172441.1 GI:14597553
KEYWORDS
SOURCE      synthetic construct.
ORGANISM      synthetic construct.
REFERENCE      1 (bases 1 to 4832)
AUTHORS      de Both,G. and de Beuckeleer,M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL      Patent: WO 0141558-A 2 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES
Source      Location/Qualifiers
1..4832
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="n-DNA of plasmid pTHW118"
misc_feature      1883..4065
/note="HpaI restriction fragment"
BASE COUNT      1528 a      883 c      932 g      1488 t
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 4602 ATAGGTGGGAGGCTATTGG 4582

RESULT 5
A60108/c      A60108      4946 bp      DNA      PAT      06-MAR-1998
LOCUS      Sequence 1 from Patent WO9706267.
DEFINITION      A60108
ACCESSION      A60108
VERSION      A60108.1 GI:3715124
SOURCE      common tobacco.
ORGANISM      Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 3266)
Marians,C., Leemans,J., De Greef,W. and De Beuckeleer,M.
TITLE      Plants with modified stamen cells
JOURNAL      Patent: EP 0344029-A 9 29-NOV-1989;
```

## KEYWORDS

SOURCE Transformation vector pTHW107.  
ORGANISM Transformation vector pTHW107  
artificial sequence; vectors.

REFERENCE 1 (bases 1 to 4946)

AUTHORS De,B.M.

TITLE Genetic transformation using a PARP inhibitor

JOURNAL Patent: WO 9706267-A 1 20-FEB-1997;

PLANT GENETIC SYSTEMS NV (BE)

Location/Qualifiers

FEATURES

source  
1. .4946  
/organism="transformation vector pTHW107"

Db 4716 ATAGGTGGAGGCTATTGG 4696

BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 4946;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTGGAGGCTATTGG 4696

## RESULT 6

A76915/c

LOCUS

DEFINITION A76915 4946 bp DNA

ACCESSION Sequence 1 from Patent EP0757102.

VERSION A76915.1 GI:6088712

KEYWORDS

SOURCE Transformation vector pTHW107.

ORGANISM Transformation vector pTHW107

artificial sequence; vectors.

REFERENCE 1 (bases 1 to 4946)

AUTHORS De,B.M.

TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR

JOURNAL Patent: EP 0757102-A 1 05-FEB-1997;

PLANT GENETIC SYSTEMS NV (BE)

Location/Qualifiers

FEATURES

source  
1. .4946  
/organism="transformation vector pTHW107"

Db 4716 ATAGGTGGAGGCTATTGG 4696

BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 4946;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTGGAGGCTATTGG 4696

## RESULT 7

AR098307/c

LOCUS

DEFINITION AR098307 4946 bp DNA

ACCESSION Sequence 1 from patent US 6074876.

VERSION AR098307.1 GI:12807564

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 4946)

AUTHORS De Block,M.

TITLE Genetic transformation using a PARP inhibitor

JOURNAL Patent: US 6074876-A 1 13-JUN-2000;

Location/Qualifiers

FEATURES

source

1. .4946

/organism="unknown"

BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 4946;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTGGAGGCTATTGG 4696

## RESULT 8

AX172440/c

LOCUS

DEFINITION AX172440 4946 bp DNA

ACCESSION Sequence 1 from Patent WO0141558.

VERSION AX172440.1 GI:14597552

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

artificial sequence.

REFERENCE 1 (bases 1 to 4946)

AUTHORS de Both,G. and de Beuckeleer,M.

TITLE Hybrid winter oilseed rape and methods for producing same

JOURNAL Patent: WO 0141558-A 1 14-JUN-2001;

Aventis CropScience N.V. (BE)

Location/Qualifiers

FEATURES

source

1. .4946

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="T-DNA of plasmid pTHW107"

misc\_feature 964. .4906

BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 4946;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTGGAGGCTATTGG 4696

## RESULT 9

AX127748/c

LOCUS

DEFINITION AX127748 5865 bp DNA

ACCESSION Sequence 1 from Patent WO0131042.

VERSION AX127748.1 GI:14134395

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

artificial sequence.

REFERENCE 1 (bases 1 to 5865)

AUTHORS Weston,B. and de Beuckeleer,M.

TITLE Male-sterile brassica plants and methods for producing same

JOURNAL Patent: WO 0131042-A 1 03-MAY-2001;

Aventis CropScience N.V. (BE)

Location/Qualifiers

FEATURES

source

1. .5865

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="T-DNA of plasmid pCOL13"

BASE COUNT 1849 a 1095 c 1149 g 1772 t

ORIGIN

```

Query Match      100.0%; Score 21; DB 6; Length 5865;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atagggtggaggctattgg 21
|||||
Db 4717 ATAGGCTGGAGGCTATTGG 4697

RESULT 10
NTTA29      NTTA29      6254 bp      DNA      PLN      09-MAY-1995
LOCUS      Tobacco anther-specific gene TA-29 and stem-specific gene TSJT1.
DEFINITION X52283
ACCESSION  X52283
VERSION    X52283.1 GI:20035
KEYWORDS   tissue specific gene.
SOURCE     common tobacco.
ORGANISM   Nicotiana tabacum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE  1 (bases 1 to 6254)
AUTHORS    Mariani T.
TITLE      Direct Submission
JOURNAL    Submitted (23-MAR-1990) Mariani T., Plant Genetic Systems, J
            Plateaustraat 22, B 9000 Gent, Belgium
REFERENCE  2 (bases 1 to 6254)
AUTHORS    Seurinck, J., Truettner, J. and Goldberg, R.B.
TITLE      The nucleotide sequence of an anther-specific gene
JOURNAL    Nucleic Acids Res. 18 (11), 3403 (1990)
MEDLINE    9028727
COMMENT    Data kindly reviewed (13-DEC-1990) by Mariani C.
FEATURES   source
            1..6254
                /organism="Nicotiana tabacum"
                /cultivar="Samsun"
                /db_xref="taxon:4097"
                /clone_lib="lambda Charon32"
                1446..1452
                /note="TSJT1"
                1477..2714
                /gene="TA29"
                /gene="TA29"
                /note="anther-specific"
                1527..2492
                /gene="TA29"
                /note="anther-specific"
                /codon_start=1
                /protein_id="CAA36524.1"
                /db_xref="GI:20036"
                /db_xref="SWISS-PROT:P24804"
                /translation="MVAPKWFVIFSMILLSLAICSGQPVTSDAIKAKADHDNLKAHT
                LSNIDAKFGGGGFGIGGWAGGGGSDTPNGYNPCSIHGCVPGFGFL
                PKPVFGVPVSPGGYVCPADIPGTGMTESKITGTSOSARLYRCKPGPMCDSDQNE
                LLLHFVPMQDKHOKHLRYGRRGIGLTVGGVGFGIGFAGWGGGGGGGSDA
                PGCSNDGCDPGFCPGCGYACAPANNPSSGGTFRHISGLSRFDGPIRCRPMCESED
                NELLHFVSPQKHENRHDHIVERSEEAHQKHDEDIIN"
                1527..2489
                /gene="TA29"
                /product="anther-specific"
                complement(<3022..>4678)
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(<3022..3075)
                /gene="TSJT1"
                /number=3
                complement(join(<3022..3075,4047..4287,4473..>4678))
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(3022..4678)
                /gene="TSJT1"

mat_peptide
            1527..2489
                /gene="TA29"
                /product="anther-specific"
                complement(<3022..>4678)
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(<3022..3075)
                /gene="TSJT1"
                /number=3
                complement(join(<3022..3075,4047..4287,4473..>4678))
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(3022..4678)
                /gene="TSJT1"

prim_transcript
            1527..2489
                /gene="TA29"
                /product="anther-specific"
                complement(<3022..>4678)
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(<3022..3075)
                /gene="TSJT1"
                /number=3
                complement(join(<3022..3075,4047..4287,4473..>4678))
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(3022..4678)
                /gene="TSJT1"

exon
            1527..2489
                /gene="TA29"
                /product="anther-specific"
                complement(<3022..>4678)
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(<3022..3075)
                /gene="TSJT1"
                /number=3
                complement(join(<3022..3075,4047..4287,4473..>4678))
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(3022..4678)
                /gene="TSJT1"

mRNA
            1527..2489
                /gene="TA29"
                /product="anther-specific"
                complement(<3022..>4678)
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(<3022..3075)
                /gene="TSJT1"
                /number=3
                complement(join(<3022..3075,4047..4287,4473..>4678))
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(3022..4678)
                /gene="TSJT1"

gene
            1527..2489
                /gene="TA29"
                /product="anther-specific"
                complement(<3022..>4678)
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(<3022..3075)
                /gene="TSJT1"
                /number=3
                complement(join(<3022..3075,4047..4287,4473..>4678))
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(3022..4678)
                /gene="TSJT1"

CDS
            1527..2489
                /gene="TA29"
                /product="anther-specific"
                complement(<3022..>4678)
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(<3022..3075)
                /gene="TSJT1"
                /number=3
                complement(join(<3022..3075,4047..4287,4473..>4678))
                /gene="TSJT1"
                /note="stem specific, weakly expressed in other organs"
                complement(3022..4678)
                /gene="TSJT1"

Query Match      100.0%; Score 21; DB 8; Length 6254;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atagggtggaggctattgg 21
|||||
Db 178 ATAGGCTGGAGGCTATTGG 198

RESULT 11
XLU70985/c
LOCUS      XLU70985
DEFINITION Xenopus laevis transforming growth factor-beta 5 gene, 5' flanking
            sequence.
ACCESSION  U70985
VERSION    U70985.1 GI:2267223
KEYWORDS   African clawed frog.
SOURCE     Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 3243)
AUTHORS    Vempati,U.D. and Kondaiah,P.
TITLE      Characterization of the 5' flanking region of the Xenopus laevis
            transforming growth factor-beta 5 (TGF-beta 5) gene
JOURNAL    Gene 208 (2), 323-329 (1998)
MEDLINE    98201629
REFERENCE  2 (bases 1 to 3243)
AUTHORS    Vempati,U.D. and Kondaiah,P.
TITLE      Direct Submission
JOURNAL    Submitted (17-SEP-1996) Center for Reproductive Biology and
            Molecular Endocrinology, Indian Institute of Science, Bangalore,
            Karnataka 560 012, India
COMMENT    On Jul 21, 1997 this sequence version replaced gi:1809300.
FEATURES   source
            1..3243
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"
            1..2890
                /note="5' flanking sequence of transforming growth
                factor-beta 5"

BASE COUNT  2042 a 1082 c 1120 g 2010 t
ORIGIN
1 atagggtggaggctattgg 21
|||||
178 ATAGGCTGGAGGCTATTGG 198

Query Match      100.0%; Score 21; DB 8; Length 6254;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atagggtggaggctattgg 21
|||||
Db 178 ATAGGCTGGAGGCTATTGG 198

RESULT 11
XLU70985/c
LOCUS      XLU70985
DEFINITION Xenopus laevis transforming growth factor-beta 5 gene, 5' flanking
            sequence.
ACCESSION  U70985
VERSION    U70985.1 GI:2267223
KEYWORDS   African clawed frog.
SOURCE     Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 3243)
AUTHORS    Vempati,U.D. and Kondaiah,P.
TITLE      Characterization of the 5' flanking region of the Xenopus laevis
            transforming growth factor-beta 5 (TGF-beta 5) gene
JOURNAL    Gene 208 (2), 323-329 (1998)
MEDLINE    98201629
REFERENCE  2 (bases 1 to 3243)
AUTHORS    Vempati,U.D. and Kondaiah,P.
TITLE      Direct Submission
JOURNAL    Submitted (17-SEP-1996) Center for Reproductive Biology and
            Molecular Endocrinology, Indian Institute of Science, Bangalore,
            Karnataka 560 012, India
COMMENT    On Jul 21, 1997 this sequence version replaced gi:1809300.
FEATURES   source
            1..3243
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"
            1..2890
                /note="5' flanking sequence of transforming growth
                factor-beta 5"

```

```
exon      <1..3224
          /number=1
CDS       2891..>3224
          /codon_start=1
          /product="transforming growth factor-beta 5 (TGF-beta 5)"
          /protein_id="AACG0371.1"
          /db_xref="GI:2281072"
          /translation="MEVLWMLLVLLVHLSSLSLSTCKAVDMEEVVRKRRIEAIKQ
ILSKLKDTPVDSEKMTVPSEAFILYNSTLEVIKATREEHVGHQNIQDIYAK
QVYRFESIT"
          3225..>3243
intron    /number=1
          970 a 675 c 676 g 922 t

BASE COUNT
ORIGIN

Query Match      92.4%; Score 19.4; DB 5; Length 3243;
Best Local Similarity 95.2%; Pred. No. 3.6;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
|||||
Db 1731 ATAGGGTGGGAGGTATATGG 1711

RESULT 12
AC022563/c
LOCUS      AC022563 149710 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-3L23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC022563
VERSION    AC022563.2 GI:9121084
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 149710)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivart,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 13, 2000 this sequence version replaced gi:6910806.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L2771
            Center clone name: 3_L_23
```





\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be processed.

\* 1 58789; contig of 58789 bp in length  
 \* 58790 58889; gap of 100 bp  
 \* 58890 67134; contig of 8245 bp in length  
 \* 67135 67234; gap of 100 bp  
 \* 67235 91255; contig of 24021 bp in length  
 \* 91256 91355; gap of 100 bp  
 \* 91356 179183; contig of 87828 bp in length.

## FEATURES

Location/Qualifiers

1..179183  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="15"  
 /map="15"  
 /clone="RP11-424I19"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 48898 a 44916 c 42628 g 42401 t 340 others

## BASE COUNT

ORIGIN

Query Match 92.4%; Score 19.4; DB 2; Length 179183;  
 Best Local Similarity 95.2%; Pred. No. 4.8;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 atagggtggggcattttg 21

|||||

Db 97951 ACAGGTGGAGCTATTGG 97931

## RESULT 14

AC073374/c

LOCUS

AC073374 167409 bp DNA HTG 23-SEP-2000  
 Homo sapiens chromosome 2 clone RP11-44108 map 2, WORKING DRAFT

SEQUENCE, 23 unordered pieces.

AC073374

AC073374.3 GI:10280771

HTG; HTGS\_PRAISE1; HTGS\_DRAFT.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 167409)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 167409)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,  
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McElrath,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 McCarthy,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

## TITLE

JOURNAL

Submitted (15-JUN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 23, 2000 this sequence version replaced gi:9857543.

All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10374

Center clone name: 441\_O8

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 154202 bases at least Q40

Consensus quality: 160990 bases at least Q30

Consensus quality: 163533 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 165209; sum-of-contigs

Quality coverage: 4.2 in Q20 bases; agarose-fp

Quality coverage: 4.3 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 23 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 11391: contig of 11391 bp in length  
 \* 11392 11491: gap of 100 bp  
 \* 11492 13778: contig of 2287 bp in length  
 \* 13779 13878: gap of 100 bp  
 \* 13879 15678: contig of 1800 bp in length  
 \* 15679 15778: gap of 100 bp  
 \* 15779 17278: contig of 1500 bp in length  
 \* 17279 17378: gap of 100 bp  
 \* 17379 19365: contig of 1987 bp in length  
 \* 19366 19465: gap of 100 bp  
 \* 19466 22846: contig of 3381 bp in length  
 \* 22847 22946: gap of 100 bp  
 \* 22947 26400: contig of 3454 bp in length  
 \* 26401 26500: gap of 100 bp  
 \* 26501 28910: contig of 2410 bp in length  
 \* 28911 29010: gap of 100 bp  
 \* 29011 51360: contig of 22350 bp in length  
 \* 51361 51460: gap of 100 bp  
 \* 51461 55176: contig of 3716 bp in length  
 \* 55177 55276: gap of 100 bp  
 \* 55277 58304: contig of 3028 bp in length  
 \* 58305 58404: gap of 100 bp  
 \* 58405 62806: contig of 4402 bp in length  
 \* 62807 62906: gap of 100 bp  
 \* 62907 69285: contig of 6379 bp in length  
 \* 69286 69385: gap of 100 bp  
 \* 69386 79312: contig of 9827 bp in length  
 \* 79313 79312: gap of 100 bp  
 \* 79313 87095: contig of 7783 bp in length  
 \* 87096 87195: gap of 100 bp  
 \* 87196 95331: contig of 8136 bp in length  
 \* 95332 95431: gap of 100 bp  
 \* 95432 102696: contig of 7265 bp in length  
 \* 102697 102796: gap of 100 bp  
 \* 102797 111775: contig of 8979 bp in length  
 \* 111776 111875: gap of 100 bp  
 \* 111876 122023: contig of 10148 bp in length  
 \* 122024 122123: gap of 100 bp  
 \* 122124 132655: contig of 10532 bp in length  
 \* 132656 132755: gap of 100 bp  
 \* 132756 145981: contig of 13225 bp in length  
 \* 145981 146080: gap of 100 bp

\* 146081 163677: contig of 17597 bp in length  
 \* 163678 163777: gap of 100 bp  
 \* 163778 167409: contig of 3632 bp in length.

## FEATURES

## Source

1. 167409  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-44108"  
 /clone\_lib="RPC1-11 Human Male BAC"

misc\_feature  
 1..11391  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left"  
 11492..11378  
 /note="assembly\_fragment"  
 misc\_feature  
 13879..15678  
 /note="assembly\_fragment"  
 15779..17278  
 /note="assembly\_fragment"  
 misc\_feature  
 17379..19365  
 /note="assembly\_fragment"  
 misc\_feature  
 19466..22846  
 /note="assembly\_fragment"  
 misc\_feature  
 22947..26400  
 /note="assembly\_fragment"  
 misc\_feature  
 26501..28910  
 /note="assembly\_fragment"  
 misc\_feature  
 29011..51360  
 /note="assembly\_fragment"  
 misc\_feature  
 51461..55176  
 /note="assembly\_fragment"  
 misc\_feature  
 55277..58304  
 /note="assembly\_fragment"  
 misc\_feature  
 58405..62806  
 /note="assembly\_fragment"  
 misc\_feature  
 62907..65285  
 /note="assembly\_fragment"  
 misc\_feature  
 69386..79212  
 /note="assembly\_fragment"  
 misc\_feature  
 79313..87095  
 /note="assembly\_fragment"  
 misc\_feature  
 87196..95331  
 /note="assembly\_fragment"  
 misc\_feature  
 95432..102696  
 /note="assembly\_fragment"  
 misc\_feature  
 102797..111775  
 /note="assembly\_fragment"  
 misc\_feature  
 111876..122023  
 /note="assembly\_fragment"  
 misc\_feature  
 122124..132655  
 /note="assembly\_fragment"  
 misc\_feature  
 132756..145980  
 /note="assembly\_fragment"  
 misc\_feature  
 146081..163677  
 /note="assembly\_fragment"  
 misc\_feature  
 163778..167409  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:right"  
 51934 a 30270 c 30508 g 52497 t 2200 others

BASE COUNT 51934 a 30270 c 30508 g 52497 t 2200 others  
 ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 167409;  
 Best Local Similarity 90.5%; Pred. No. 39;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21  
 ||||| ||||| ||||| |||||

Db 126062 ATAGACTTGGAGGCTATTTCG 126042

## RESULT 15

## AC013820

## LOCUS

## DEFINITION

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## RESULT 15

## AC013820

## LOCUS

## DEFINITION

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## RESULT 15

## AC013820

## LOCUS

## DEFINITION

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## RESULT 15

## AC013820

## LOCUS

## DEFINITION

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

## AC013820

```
* 27354 41791: contig of 14438 bp in length
* 41792 41891: gap of 100 bp
* 41892 61037: contig of 19146 bp in length
* 61038 61137: gap of 100 bp
* 61138 88113: contig of 26976 bp in length
* 88114 88213: gap of 100 bp
* 88214 131619: contig of 43406 bp in length
* 131620 131719: gap of 100 bp
* 131720 179510: contig of 47791 bp in length.
```

FEATURES

```
Source
1..179510
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-21P24"
/clone_lib="RPC1-11 Human Male BAC"
1..2601
/note="assembly_fragment"
misc_feature 2702..8297
/note="assembly_fragment"
misc_feature 8398..27253
/note="assembly_fragment"
misc_feature 27354..41791
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
misc_feature 41892..61037
/note="assembly_fragment"
misc_feature 61138..88113
/note="assembly_fragment"
misc_feature 88214..131619
/note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 131720..179510
/note="assembly_fragment"
BASE COUNT 59127 a 32374 c 31805 g 55501 t 703 others
ORIGIN
```

```
Query Match 84.8%; Score 17.8; DB 2; Length 179510;
Best Local Similarity 90.5%; Pred.No.39;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atagggtggaggctatttgg 21
|||||
Db 107231 ATAGAGTTGGAGGCTATTGG 107251
```

Search completed: February 25, 2002, 18:00:06  
Job time: 18414 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:19 ; Search time 716.55 Seconds  
(without alignments)  
25.126 Million cell updates/sec

Title: US-09-698-903B-3  
Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	21	100.0	21	AAD06992	PCR primer MLD008
2	21	100.0	3265	AA02188	TA29 gene Insert 1
3	21	100.0	4832	AAH25423	Nucleotide sequenc
4	21	100.0	4946	AAT59531	T-DNA of plasmid p
5	21	100.0	4946	AAH25422	Nucleotide sequenc
6	21	100.0	5864	AAT39339	Plasmid pTCO113 T-
7	21	100.0	5865	AAD06990	Chimeric T-DNA of
8	16.8	80.0	1983	AAH83820	Dengue-2 virus der
9	16.8	80.0	10723	AAH49304	cDNA encoding poly
10	16.8	80.0	10723	AAH49303	cDNA sequence enco
11	16.4	78.1	694	AAH68287	Human lung tumour

c 12	16.4	78.1	857	14	AAQ46850	Recombinant human
c 13	16.4	78.1	857	17	AAH14564	Human milk kappa-c
c 14	16.4	78.1	858	16	AAQ89598	Human kappa-casein
c 15	16.4	78.1	2185	20	AAZ21101	Human secreted pro
c 16	16.4	78.1	13104	14	AAQ46852	Clone of recombina
c 17	16.2	77.1	1418	20	AAQ4867	rodent DCMPI C-lec
c 18	16.2	77.1	24978	20	AAQ60209	Genomic DNA sequen
c 19	16.2	77.1	24979	21	AAAS2321	SEQ ID 3 of US5914
c 20	15.8	75.2	394	21	AAAC03474	Human secreted pro
c 21	15.8	75.2	660	21	AAAF12948	Aspergillus oryzae
c 22	15.8	75.2	777	15	AAQ70891	Tomato spotted wil
c 23	15.8	75.2	1026	16	AAQ94047	Human ALD gene exo
c 24	15.8	75.2	1549	22	AAH14886	Human cDNA sequenc
c 25	15.8	75.2	1556	22	AAH74185	s1 RNA binding reg
c 26	15.8	75.2	1577	22	AAH159214	Human polynucleoti
c 27	15.8	75.2	1591	22	AAH16816	Human cDNA sequenc
c 28	15.8	75.2	1602	22	AAH15727	Human cDNA sequenc
c 29	15.8	75.2	1643	19	AAV43711	Human sodium-depen
c 30	15.8	75.2	1644	21	AAH76111	Human ORFX ORF1666
c 31	15.8	75.2	1795	19	AAV57910	Human haemochromat
c 32	15.8	75.2	3049	15	AAQ70890	Tomato spotted wil
c 33	15.8	75.2	9139	21	AAZ35274	Soybean retroelime
c 34	15.4	73.3	2338	22	AAH16691	Human cDNA sequenc
c 35	15.4	73.3	3024	20	AAZ42247	Human normal bladd
c 36	15.4	73.3	4636	19	AAV23920	Human alpha3 integ
c 37	15.4	73.3	5769	19	AAV15004	Receptor protein t
c 38	15.4	73.3	44453	20	AAZ3519	Human kidney amino
c 39	15.2	72.4	139	12	AAQ13998	Template DNA. Syn
c 40	15.2	72.4	140	14	AAQ35327	DNA amplification
c 41	15.2	72.4	455	22	AAH12272	Probe #2205 for ge
c 42	15.2	72.4	455	22	AAH13627	Probe #2313 used t
c 43	15.2	72.4	455	22	AAH102189	Probe #2180 used t
c 44	15.2	72.4	649	21	AAH14117	Aspergillus oryzae
c 45	15.2	72.4	1302	21	AAH50359	Arabidopsis thalia

## ALIGNMENTS

RESULT 1  
AAD06992 AAD06992 standard; DNA; 21 BP.  
XX  
AC AAD06992;  
XX  
XX 06-AUG-2001 (first entry)  
XX  
DE PCR primer MLD008 to amplify a fragment of pTCO113.  
XX  
XX Plasmid pTCO113; transgenic Brassica plant; transformation event;  
KW male-sterility gene; tobacco; PCR primer; probe; ss.  
XX  
XX Chimeric - Bacillus amyloliquefaciens.  
OS Chimeric - Nicotiana tabacum.  
XX  
XX WO200131042-A2.  
XX  
PD 03-MAY-2001.  
XX  
XX 26-OCT-2000; 2000WO-EP10680.  
PF  
XX 29-OCT-1999; 99US-0430497.  
PR  
XX (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
PI Weston B, De Beuckeleer M;  
XX  
DR WPI; 2001-300517/31.  
XX  
PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX



```

FT FT /*tag= l
FT FT /note= "downstream of Bacillus amyloliquefaciens
FT FT barstar coding region"
FT CDS complement (2981..3253)
FT FT /*tag= m
FT FT /note= "Barstar gene coding region from Bacillus
FT FT amyloliquefaciens"
FT FT complement (3254..4762)
FT FT /*tag= n
FT FT /note= "another-specific gene TA29 promoter from
FT FT Nicotiana tabacum"
FT FT 4763..4807
FT FT /*tag= o
FT FT /note= "Synthetic polylinker derived sequences"
FT FT 4808..4832
FT FT /*tag= p
FT FT /note= "left border repeat from TL-DNA from pTiB6S3"
XX XX
XX PN WO200141558-A1.
XX XX
XX PD 14-JUN-2001.
XX XX
XX PF 06-DEC-2000; 2000WO-EPI2872.
XX XX
XX PR 08-DEC-1999; 99US-0457037.
XX XX
XX PA (AVET ) AVENTIS CROPS SCIENCE NV.
XX XX
XX PI De Both G, De Beuckeleer M;
XX XX
XX DR WPI; 2001-381419/40.
XX XX
XX XX Transgenic winter oilseed rape plants suited for producing hybrid seed
XX PT with improved qualities, comprises a male-sterility gene and fertility
XX PT restorer gene, integrated into the genome
XX PS
XX PS Example 1; Page 80-82; 98pp; English.
XX XX
XX CC The specification describes a pair of transgenic winter oilseed rape
XX CC plants suited for producing hybrid seed. One of the plants has an
XX CC expression cassette comprising a male-sterility gene, and the other
XX CC plant has an expression cassette comprising a fertility restorer gene,
XX CC integrated into the genome. The fertility restorer gene is capable of
XX CC preventing the activity of the male-sterility gene. The plant pair is
XX CC useful for producing hybrid seed. Plants developed from the hybrid
XX CC seed have agronomic performance, genetic stability and adaptability to
XX CC different genetic backgrounds. The present sequence represents
XX CC plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
XX CC a fertility restorer gene. The plasmid is used to create transgenic
XX CC plants of the invention.
XX SQ Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;

Query Match 100.0%; Score 21; DB 22; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atagggtggaggctatttgg 21
| | | | | | | | | | | | | | | | | | | | |
Db 4602 ATAGGGTGGAGGCTATTGG 4582

RESULT 4
AAT59531/c
ID AAT59531 standard; DNA; 4946 BP.
XX AC
XX AC AAT59531;
XX XX
XX DT 07-MAY-1997 (first entry)
XX DE T-DNA of plasmid pTHW107.
XX XX

```

```

KW KW Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PARP;
KW KW niacinamide; Agrobacterium; T-DNA; male sterile; barnase;
KW KW ribonuclease; RNase; cereal; wheat; oilseed rape; Brassica napus;
XX XX plasmid pTHW107; ds.
OS OS Chimeric Agrobacterium sp.;
OS OS Chimeric Arabidopsis thaliana;
XX XX Chimeric Nicotiana tabacum.
XX FT Key Location/Qualifiers
FT FT misc_RNA complement (1..25)
FT FT /*tag= a
FT FT /label= RB
FT FT /note= "T-DNA right border"
FT 3'UTR complement (97..330)
FT FT /*tag= b
FT FT /label= 3'g7
FT FT /note= "3' untranslated region contg. the poly-A
FT FT signal of gene-7 of Agrobacterium T-DNA"
FT CDS complement (331..882)
FT FT /*tag= c
FT FT /label= Bar
FT FT /product= phosphinothricin acetyltransferase
FT FT complement (883..2608)
FT FT /*tag= d
FT FT /label= PSSU
FT FT /note= "promoter region of Rubisco small subunit
FT FT gene of Arabidopsis thaliana"
FT 3'UTR complement (2658..3031)
FT FT /*tag= e
FT FT /label= 3'nos
FT FT /note= "3' untranslated region contg. the poly-A
FT FT signal of the nopaline-synthase gene of
FT FT Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT FT /*tag= f
FT FT /label= Barnase
FT FT /product= barnase
FT FT complement (3368..4876)
FT FT /*tag= g
FT FT /label= pTA29
FT FT /note= "promoter region of tobacco TA29 gene"
FT FT complement (4822..4946)
FT FT /*tag= h
FT FT /label= LB
FT FT /note= "T-DNA left border"
XX XX EP757102-A1.
XX PN
XX XX 05-FEB-1997.
XX PD
XX XX 04-AUG-1995; 95EP-0401844.
XX PF
XX XX 04-AUG-1995; 95EP-0401844.
XX PR
XX XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PA
XX XX De Block M;
XX PI
XX XX WPI; 1997-111050/11.
XX DR
XX XX Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase
XX PT inhibitor - reduces the cultured cells response to stress and
XX PT reduces metabolism
XX FT
XX PS Example 3; Page 13-16; 25pp; English.
XX XX
XX CC Plasmid pTHW107 is a vector carrying T-DNA (AAT59531) comprising a
XX CC barnase coding sequence under control of the tobacco TA29 gene
XX CC stamen-specific promoter and a phosphinothricin acetyltransferase
XX CC coding sequence under control of an Arabidopsis Rubisco small
XX CC subunit gene promoter. Oilseed rape hypocotyl explants were
XX CC infected with Agrobacterium tumefaciens C58C1Rif carrying vector

```

CC pTHW107 and helper T1 plasmid pMP60. In some treatments, the  
 CC hypocotyls were treated with the poly-(ADP-ribose) polymerase  
 CC inhibitor niacinamide (250 mg/l) 4 days prior to infection.  
 CC Plants regenerated from niacinamide-treated transformed calli  
 CC had a low copy number and displayed less variation in the  
 CC expression profile of the transgenes.  
 XX  
 SQ Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match 100.0%; Score 21; DB 18; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.42; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0;

OY 1 atagggtggaggctatttgg 21  
 |||||  
 Db 4716 ATAGGGTGGGAGGCTATTGG 4696

RESULT 5  
 AAH25422/c  
 ID AAH25422 standard; DNA; 4946 BP.  
 XX  
 AC AAH25422;  
 XX

DT 22-AUG-2001 (first entry)

XX Nucleotide sequence of plasmid pTHW107.

DE Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; ss.  
 XX

OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 OS Arabidopsis thaliana.  
 OS Bacillus amyloliquefaciens.  
 OS Nicotiana tabacum.

FH Key Location/Qualifiers

FT misc\_feature 1..25  
 /tag= a  
 /note= "right border repeat from TL-DNA from pTiB6S3"  
 FT misc\_feature 26..97  
 /tag= b  
 /note= "synthetic polylinker derived sequences"  
 FT 3'UTR complement (98..309)  
 /tag= c  
 /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"  
 FT misc\_feature 310..330  
 /tag= d  
 /note= "synthetic polylinker derived sequences"  
 FT CDS 331..882  
 /tag= e  
 /note= "Streptomyces hygroscopicus bar gene"  
 FT promoter complement (883..2608)  
 /tag= f  
 /note= "AtSIA ribulose-1,5-biphosphate carboxylase  
 small subunit gene from Arabidopsis thaliana"  
 FT misc\_feature 2609..2658  
 /tag= g  
 /note= "synthetic polylinker derived sequences"  
 FT 3'UTR complement (2659..2919)  
 /tag= h  
 /note= "tagl fragment from 3' UTR of nopaline  
 synthase gene from T-DNA of pTiT37 and  
 containing plant polyadenylation signals"

FT 3'UTR 2920..3031  
 /tag= i  
 /note= "3'UTR downstream of Bacillus amyloliquefaciens  
 barnase coding region"  
 FT CDS complement (3032..3367)  
 /tag= j  
 /note= "Barnase coding region from Bacillus  
 .  
 .

FT promoter complement amyloliquefaciens"  
 /tag= k  
 /note= "anther-specific gene TA29 promoter from  
 Nicotiana tabacum"  
 FT misc\_feature 4878..4921  
 /tag= l  
 /note= "synthetic polylinker derived sequences"  
 FT misc\_feature 4922..4946  
 /tag= m  
 /note= "left border repeat from TL-DNA from pTiB6S3"

WO200141558-A1.

14-JUN-2001.

06-DEC-2000; 2000WO-EP12872.

08-DEC-1999; 99US-0457037.

(AVET ) AVENTIS CROPS SCIENCE NV.

De Both G, De Beuckeleer M;

WPI; 2001-381419/40.

Transgenic winter oilseed rape plants suited for producing hybrid seed  
 with improved qualities, comprises a male-sterility gene and fertility  
 restorer gene, integrated into the genome -

Example 1; Page 78-80; 98pp; English.

The specification describes a pair of transgenic winter oilseed rape  
 plants suited for producing hybrid seed. One of the plants has an  
 expression cassette comprising a male-sterility gene, and the other  
 plant has an expression cassette comprising a fertility restorer gene,  
 integrated into the genome. The fertility restorer gene is capable of  
 preventing the activity of the male-sterility gene. The plant pair is  
 useful for producing hybrid seed. Plants developed from the hybrid  
 seed have agronomic performance, genetic stability and adaptability to  
 different genetic backgrounds. The present sequence represents  
 plasmid pTHW107. This plasmid comprises the barnase gene, which acts as  
 a male-sterility gene. The plasmid is used to create transgenic plants  
 of the invention.

Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.42;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atagggtggaggctatttgg 21  
 |||||

Db 4716 ATAGGGTGGGAGGCTATTGG 4696

RESULT 6  
 AAT39339/c  
 ID AAT39339 standard; DNA; 5864 BP.  
 XX  
 AC AAT39339;  
 XX

DT 22-JAN-1997 (first entry)

DE Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.

KW Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;

XX transgenic plant; oilseed rape; canole; Brassica napus; ds.

OS Synthetic.

FH Key Location/Qualifiers



```
FT misc_feature complement (1..25)
FT /tag= a
FT /label= RB
FT /note= "right border of Agrobacterium T-DNA"
FT polyA_signal complement (98..330)
FT /tag= b
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT CDS complement (331..882)
FT /tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin
FT acetyltransferase"
FT promoter complement (883..2608)
FT /tag= d
FT /label= Psu
FT /note= "promoter of Arabidopsis Rubisco small
FT subunit gene"
FT polyA_signal complement (2659..3031)
FT /tag= e
FT /label= 3'nos
FT /note= "region containing polyA signal of nopaline
FT synthase gene of Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /tag= f
FT /label= Barnase
FT /note= "Bacillus amyloliquefaciens barnase coding
FT region"
FT promoter complement (3368..4877)
FT /tag= g
FT /label= PTA29
FT /note= "promoter of stamen-specific TA29 gene of
FT Nicotiana tabacum"
FT promoter 4924..5216
FT /tag= h
FT /label= Pnos
FT /note= "promoter of nopaline synthase gene of
FT Agrobacterium T-DNA"
FT CDS 5217..5489
FT /tag= i
FT /label= Barstar
FT /note= "region coding for barstar of Bacillus
FT amyloliquefaciens"
FT polyA_signal 5490..5765
FT /tag= j
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT misc_feature complement (5840..5864)
FT /tag= k
FT /label= LB
FT /note= "left border of Agrobacterium T-DNA"
FT
FT
FT WO9626283-A1.
FT
FT
FT 29-AUG-1996. 96WO-EP00722.
FT
FT 21-FEB-1996; 96WO-EP00722.
FT
FT 21-FEB-1995; 95EP-0400364.
FT
FT (PLBZ ) PLANT GENETIC SYSTEMS NV.
FT
FT Botterman J, Cornelissen M, Michiels F;
FT WPI; 1996-402373/40.
FT
FT Prodn. of male sterile plants by transforming with a chimaeric
FT construct - comprising a male sterility DNA e.g. barnase and a
FT co-regulating gene, e.g. barstar, into the nuclear genome, useful
FT for generating hybrid cultivars
FT
```

```
PS Example 3; Page 33-3743-47; 56pp; English.
XX
XX Plasmid pTCO113 (AAT39339) is a T-DNA vector containing a bar gene
XX under control of the PSSU promoter, a barnase gene under control
XX of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
XX gene under control of the Pnos promoter. 87% Of oilseed rape
XX plants regenerated after Agrobacterium-mediated transformation
XX using pTCO113 were male sterile. Barnase expression disturbed the
XX function of stamen cells leading to male sterility. Constitutive
XX expression of barstar counteracted any low level expression of
XX barnase in non-stamen tissue.
XX
XX Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;
SQ
Query Match 100.0%; Score 21; DB 17; Length 5864;
Best Local Similarity 100.0%; Pred. NO. 0.43;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atagggtggaggctatttgg 21
Db 4716 ATAGGTTGGGAGGCTATTGG 4696
RESULT 7
AAD06990/c
ID AAD06990 standard; DNA; 5865 BP.
XX
AC AAD06990;
XX
XX 06-AUG-2001 (first entry)
XX
XX Chimeric T-DNA of plasmid pTCO113.
XX
XX T-DNA; plasmid pTCO113; transgenic Brassica plant; transformation event;
XX male-sterility gene; chimeric; tobacco; ds.
XX
XX Chimeric - Streptomyces hygroscopicus.
XX Chimeric - Arabidopsis thaliana.
XX Chimeric - Bacillus amyloliquefaciens.
XX Chimeric - Nicotiana tabacum.
XX Chimeric - Agrobacterium tumefaciens.
XX Chimeric - Unidentified.
XX
XX Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "Right border repeat from the TL-DNA from
FT pTiB6S3"
FT misc_feature 26..53
FT /tag= b
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 54..90
FT /tag= c
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 3'UTR 98..309
FT /tag= d
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT misc_feature 310..331
FT /tag= e
FT /note= "Synthetic polylinker derived sequence"
FT CDS 332..883
FT /tag= f
FT /product= "Protein encoded by bialaphos resistance
FT gene (bar) of Streptomyces hygroscopicus"
FT promoter 884..2609
FT /tag= g
FT /note= "Promoter from the atSIA ribulose-1,5-biphosphate
FT carboxylase small subunit gene from Arabidopsis thaliana."
FT misc_feature 2610..2659
FT /tag= h
FT
```



PT Chimeric live, infectious, attenuated yellow fever viruses used for  
PT preventing and treating diseases caused by flaviviruses have prM-E  
PT nucleotide sequence from a second, different flavivirus as functional  
PT yellow fever prM-E is not expressed -  
XX  
PS Disclosure; Page 196-198; 232pp; English.  
XX  
CC The invention relates to a chimeric live, infectious, attenuated virus  
CC comprising a yellow fever virus with the nucleotide sequence encoding a  
CC prM-E protein deleted, truncated or mutated so that functional yellow  
CC fever virus prM-E protein is not expressed and also integrated into the  
CC genome of the yellow fever virus a nucleotide sequence encoding a prM-E  
CC protein of a second, different flavivirus so that the prM-E protein of  
CC the second flavivirus is expressed. The chimeric live, infectious,  
CC attenuated virus is used to prepare medicaments for preventing or  
CC treating flavivirus infection in a patient. The yellow fever virus vector  
CC produces its gene product (tumor antigen or cytokine) in cells of the  
CC lymphoid or reticuloendothelial system or in a precursor of these systems  
CC in patients with cancer. Flaviviruses replicate in the cytoplasm of cells  
CC so that the virus replication does not involve integration of the viral  
CC genome into the host cell. The present sequence represents a DNA fragment  
CC derived from dengue-2 virus.  
XX  
SQ Sequence 1983 BP; 645 A; 409 C; 509 G; 420 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 1983;  
Best Local Similarity 90.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 taggggtgggaggtatttgg 21  
||||||| ||||| |||||  
Db 666 TAGGGTGGCAGCGCTGTTGG 647

RESULT 9  
AAT49304/C  
ID AAT49304 standard; cDNA; 10723 BP.  
XX  
AC AAT49304;  
XX  
DT 12-SEP-1997 (first entry)  
XX  
DE cDNA encoding polyprotein of attenuated DEN-2 virus, PDK-53.  
XX  
KW Dengue 2 virus; polyprotein; capsid; prM; M; E; NS1; NS2A; NS2B; NS3;  
KW NS4A; NS4B; NS5; PDK-53; quadravalent vaccine; immunity; serotype;  
KW chimeric DEN-2/1 virus; chimeric DEN-2/3 virus; chimeric DEN-2/4 virus;  
KW dengue fever; fatal dengue haemorrhagic fever; dengue shock syndrome;  
KW DHF; DSS; ss.  
XX  
OS Dengue 2 virus, strain 16681.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT CDS 97..10272  
FT /\*tag= a  
FT /product= DEN-2 attenuated polyprotein  
FT /transl\_except(pos:643..645,aa:Xaa)  
FT /transl\_except(pos:1135..1137,aa:Xaa)  
FT /transl\_except(pos:1139..1195,aa:Xaa)  
FT /transl\_except(pos:2809..2811,aa:Xaa)  
FT /transl\_except(pos:3040..3042,aa:Xaa)  
FT /transl\_except(pos:9208..9210,aa:Xaa)  
FT /note= "Xaa = unknown amino acid"  
FT 57  
FT mutation  
FT /\*tag= b  
FT /note= "C>T mutation"  
FT 524  
FT mutation  
FT /\*tag= C  
FT /note= "A>T mutation, causes Asp to Val substitution"  
FT 2055  
FT mutation  
FT /\*tag= d

FT mutation  
FT /note= "C>T mutation"  
FT 2579  
FT /\*tag= e  
FT /note= "G>A mutation, causes Gly to Asp substitution"  
FT 4018  
FT /\*tag= f  
FT /note= "C>T mutation, causes Leu to Phe substitution"  
FT 5547  
FT /\*tag= g  
FT /note= "C>T mutation"  
FT 6599  
FT /\*tag= h  
FT /note= "G>C mutation, causes Gly to Ala substitution"  
FT 8571  
FT /\*tag= i  
FT /note= "C>T mutation"  
XX  
PN WO9640933-A1.  
XX  
XX 19-DEC-1996.  
PD  
XX  
PF 06-JUN-1996; 96WO-US09209.  
XX  
PR 07-JUN-1995; 95US-0483292.  
XX  
PA (UYMA-) UNIV MAHIDOL AT SALAYA.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Bhamarapravati N, Butrapet S, Chang J, Gubler DJ;  
PI Halstead SB, Kinney R, Trent DW;  
XX  
DR WPI; 1997-052330/05.  
DR P-PSDB; AAW06591.  
XX  
PT PDK-53, a clone of infectious attenuated Dengue 2 virus strain 16681  
PT - also chimeric DEN-2/1, DEN-2/3 and DEN-1/4 viruses, used as a  
PT quadravalent vaccine for protecting against Dengue virus infection  
XX  
PS Claim 27; Page 122-136; 261pp; English.  
XX  
CC This sequence encodes the polyprotein from an attenuated derivative  
CC of Dengue 2 virus, strain 16681. The derivative is designated PDK-53.  
CC The polyprotein comprises the capsid, prM, M, E, NS1, NS2A, NS2B, NS3,  
CC NS4A, NS4B and NS5 proteins. The PDK-53 viral sequence may be used in  
CC the production of a quadravalent vaccine which provides immunity against  
CC all four serotypes of dengue virus. The vaccine also comprises a  
CC chimeric DEN-2/1 virus, a chimeric DEN-2/3 virus, and/or a chimeric  
CC DEN-2/4 virus. The new quadravalent vaccines are used to protect  
CC against infection by all four serotypes of dengue virus, DEN-1, DEN-2,  
CC DEN-3 and DEN-4, which can lead to dengue fever or fatal dengue  
CC haemorrhagic fever/dengue shock syndrome (DHF/DSS). Host cells are  
CC used to produce the recombinant protein products of the DNA constructs  
CC which are used in the vaccines.  
XX  
SQ Sequence 10723 BP; 3548 A; 2196 C; 2711 G; 2261 T; 7 other;

Query Match 80.0%; Score 16.8; DB 18; Length 10723;  
Best Local Similarity 90.0%; Pred. No. 48;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 taggggtgggaggtatttgg 21  
||||||| ||||| |||||  
Db 1104 TAGGGTGGCAGCGCTGTTGG 1085  
RESULT 10  
AAT49303/C  
ID AAT49303 standard; cDNA; 10723 BP.  
XX  
AC AAT49303;  
XX  
DT 11-SEP-1997 (first entry)

```

XX DE cDNA sequence encoding polyprotein of DEN-2 virus, strain 16681.
XX KW Dengue 2 virus; polyprotein; capsid; prM; M; E; NS1; NS2A; NS2B; NS3;
XX KW NS4A; NS4B; NS5; pDK-53; quadravalent vaccine; immunity; serotype;
XX KW chimeric DEN-2/1 virus; chimeric DEN-2/3 virus; chimeric DEN-2/4 virus;
XX KW dengue fever; fatal dengue haemorrhagic fever; dengue shock syndrome;
XX KW DHF; DSS; ss.
XX OS Dengue 2 virus, strain 16681.
XX FH Key Location/Qualifiers
XX FT CDS 97..10272
XX FT /*tag= a
XX FT /product= DEN-2 polyprotein
XX FT /transl_except(pos:9208..9210, aa:Xaa)
XX FT /note= "Xaa = unknown amino acid"
XX PN WO9640933-A1.
XX PD 19-DEC-1996.
XX PF 06-JUN-1996; 96WO-US09209.
XX PR 07-JUN-1995; 95US-0483292.
XX PA (UYMA-) UNIV MAHIDOL AT SALAYA.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Bhamarapavati N, Butrapet S, Chang J, Gubler DJ;
XX PI Halstead SB, Kinney R, Trent DW;
XX DR WPI; 1997-052330/05.
XX PT P-PSDB; AAW06590.
XX PT POK-53, a clone of infectious attenuated Dengue 2 virus strain 16681
XX PT - also chimeric DEN-2/1, DEN-2/3 and DEN-1/4 viruses, used as a
XX PT quadravalent vaccine for protecting against Dengue virus infection
XX PS Claim 23; Page 107-121; 261pp; English.
XX CC This sequence encodes the polyprotein from Dengue 2 virus, strain 16681.
XX CC The polyprotein comprises the capsid, prM, M, E, NS1, NS2A, NS2B, NS3,
XX CC NS4A, NS4B and NS5 proteins. A clone of this wildtype viral sequence,
XX CC POK-53, may be used in the production of a quadravalent vaccine which
XX CC provides immunity against all four serotypes of dengue virus. The
XX CC vaccine also comprises a chimeric DEN-2/1 virus, a chimeric DEN-2/3
XX CC virus, and/or a chimeric DEN-2/4 virus. The new quadravalent vaccines
XX CC are used to protect against infection by all four serotypes of dengue
XX CC virus, DEN-1, DEN-2, DEN-3 and DEN-4, which can lead to dengue fever or
XX CC fatal dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS). Host
XX CC cells are used to produce the recombinant protein products of the DNA
XX CC constructs which are used in the vaccines.
XX SQ Sequence 10723 BP; 3553 A; 2200 C; 2713 G; 2255 T; 2 other;

Query Match 80.0%; Score 16.8; DB 18; Length 10723;
Best Local Similarity 90.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggggtggaggcgtattgg 21
||||| ||||| |||||
DB 1104 TAGGGTGGCAGGCTGTTG 1085

RESULT 11
AAF68287
ID AAF68287 standard; cDNA; 694 BP.
XX AC AAF68287;
XX DT 12-APR-2001 (first entry)

```

```

XX DE Human lung tumour protein related nucleotide sequence SEQ ID NO:205.
XX KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
XX KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
XX KW cytostatic; antisense inhibition; ss.
XX OS Homo sapiens.
XX PN WO200100828-A2.
XX PD 04-JAN-2001.
XX PF 30-JUN-2000; 2000WO-US18061.
XX PR 30-JUN-1999; 99US-0346492.
XX PR 15-OCT-1999; 99US-0419356.
XX PR 17-DEC-1999; 99US-0466867.
XX PR 30-DEC-1999; 99US-0476300.
XX PR 06-MAR-2000; 2000US-0519642.
XX PR 22-MAR-2000; 2000US-0533077.
XX PR 10-APR-2000; 2000US-0546259.
XX PR 27-APR-2000; 2000US-0560406.
XX PR 05-JUN-2000; 2000US-0589184.
XX PA (CORI-) CORIXA CORP.
XX PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX PI Retter MW, Mannion J;
XX DR WPI; 2001-071488/08.
XX PT Lung tumor-associated proteins and the nucleic acids that encode them,
XX PT useful for preventing, diagnosing and treating lung cancer -
XX PS Claim 4; Page 213; 436pp; English.
XX CC The present invention describes immunogenic portions of lung tumour-
XX CC associated proteins (I) and the nucleic acids (NAs) that encode them.
XX CC (I) have cytostatic activity and can be used in gene therapy, antisense
XX CC inhibition and in vaccines. The NAs and the lung tumour-associated
XX CC proteins they encode may be used in the prevention, treatment and
XX CC diagnosis of diseases associated with their inappropriate expression,
XX CC especially lung cancers. For example, the NAs may be administered to
XX CC treat diseases by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of the protein by expressing inactive proteins
XX CC or to supplement the patient's own production of (I). Additionally, the
XX CC NAs may be used to produce the lung-tumour associated protein, according
XX CC to standard recombinant DNA methodology. Conversely, antisense NA
XX CC molecules may be administered to down regulate protein expression by
XX CC binding with the cells own genes and preventing their expression. The NA
XX CC and complementary sequences may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar NA sequences in
XX CC samples, and hence which patients may be in need of treatment for lung
XX CC cancer. The (I) may be used as antigens in the production of antibodies
XX CC and in assays to identify modulators (agonists and antagonists) of the
XX CC expression and activity of the protein. AAF68083 to AAF68878 and
XX CC AAB76848 to AAB76878 represent human lung tumour protein related
XX CC nucleotide and protein sequences which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 694 BP; 202 A; 111 C; 154 G; 227 T; 0 other;

Query Match 78.1%; Score 16.4; DB 22; Length 694;
Best Local Similarity 94.4%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aggggtggaggcgtattgg 20
||||| ||||| |||||
DB 290 aggggtggaggcgtattgg 307

```

```
RESULT 12
AAQ46850/c
ID AAQ46850 standard; cDNA; 857 BP.
XX AC AAQ46850;
XX AC
XX DT 26-JAN-1994 (first entry)
XX DE Recombinant human kappa casein gene.
XX KW Casein; supplement; milk; pharmaceutical; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 45..593
XX FT /*tag= a
XX FT
XX mat_peptide /product= Human kappa casein.
XX FT 45..593
XX FT /*tag= b
XX FT sig_peptide 45..104
XX FT /*tag= c
XX PN WO9315196-A.
XX PN
XX PD 05-AUG-1993.
XX XX
XX PF 25-JAN-1993; 93WO-DK000024.
XX XX
XX PR 23-JAN-1992; 92DK-0000088.
XX XX
XX PA (SYMB-) SYMBICOM AB.
XX XX
XX PI Bergstroem S, Hansson L, Hernell O, Stroemqvist M;
XX PI Toernell J;
XX DR WPI; 1993-258675/32.
XX DR P-PSDB; AAR39351.
XX XX
XX PT DNA encoding human kappa-casein - used for obtaining recombinant
XX PT polypeptide(s) for use as nutrient supplements, partic. in infant
XX PT formulae
XX XX
XX PS Claim 14; Page 88-89; 110pp; English.
XX XX
XX CC The recombinant human kappa casein is produced in high yields by
XX CC means of either a eukaryotic or prokaryotic expression system. It
XX CC is used as a nutrient supplement in milk based products to provide a
XX CC substantial improvement of the nutritional and biological value of
XX CC the formulae, making it closer in similarity to human milk. It can
XX CC also be used as a pharmaceutical.
XX XX
XX SQ Sequence 857 BP; 307 A; 220 C; 128 G; 202 T; 0 other;

Query Match 78.1%; Score 16.4; DB 14; Length 857;
Best Local Similarity 94.4%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 gggtggggaggtatttgg 21
Db 359 GGGTGGGTGGCTATTGG 342

RESULT 13
AAT14564/c
ID AAT14564 standard; cDNA; 857 BP.
XX AC AAT14564;
XX AC
XX DT 24-OCT-1996 (first entry)
XX XX
XX DE Human milk kappa-casein cDNA.
```

```
XX Milk protein; kappa-casein; rotavirus inhibition; gastroenteritis;
KW infant formulation; immunodeficiency; diarrhoea; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 45..593
XX FT /*tag= a
XX PN WO9608269-A1.
XX PD 21-MAR-1996.
XX XX
XX PF 05-MAY-1995; 95WO-US05676.
XX PR 16-SEP-1994; 94US-0308883.
XX PR 16-SEP-1994; 94US-0308882.
XX PA (ABBO ) ABBOTT LAB.
XX PI Baxter JH, Cummings RD, Mukerji P, Prieto PA, Seo AE;
XX WPI; 1996-179724/18.
XX DR
XX XX
XX PT Kappa-casein used as anti-rotaviral infection agent in nutritional
XX PT product - to prevent, retard or treat rotavirus infection,
XX PT especially in infants, children and immuno-deficient patients
XX XX
XX PS Claim 1; Fig 8; 42pp; English.
XX CC
XX CC AAT14564 encodes human milk kappa-casein which is useful as a component
XX CC of an enteral nutritional product. Kappa-casein has anti-rotaviral
XX CC activity, it inhibits the attachment of human rotavirus to mammalian
XX CC cells. The kappa-casein is useful in the treatment and prevention of
XX CC onset of gastroenteritis and other diarrhoeal diseases caused by
XX CC rotaviridae esp. in infants, children or immunodeficient patients.
XX CC Kappa-casein appeared to have no side effects on the subjects treated
XX CC and is unlikely to be toxic or cause an allergic reaction.
XX XX
XX SQ Sequence 857 BP; 309 A; 219 C; 127 G; 202 T; 0 other;

Query Match 78.1%; Score 16.4; DB 17; Length 857;
Best Local Similarity 94.4%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 gggtggggaggtatttgg 21
Db 359 GGGTGGGTGGCTATTGG 342

RESULT 14
AAQ89598/c
ID AAQ89598 standard; cDNA; 858 BP.
XX AC AAQ89598;
XX AC
XX DT 06-NOV-1995 (first entry)
XX XX
XX DE Human kappa-casein cDNA.
XX XX
XX KW Kappa-casein; milk protein; ss.
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 107..655
XX FT /*tag= a
XX PN US5391497-A.
XX XX
XX PD 21-FEB-1995.
```

XX 13-OCT-1992; 92US-0962569.  
XX 13-OCT-1992; 92US-0962569.  
XX (COLS ) UNIV COLORADO FOUND INC.  
XX Chang Y, Ham RG, Jeffers KF, Menon RS;  
XX WPI; 1995-160470/21.  
XX P-PSDB; AAR72696.  
XX DNA encoding human kappa-casein - used for the prodn. of large amts. of  
XX highly purified kappa-casein milk protein for infant use.  
XX Claim 4; Column 13-16; 14pp; English.  
XX A commercial cDNA library prepd. in lambda gtil from mRNA obtd. from  
XX human breast tissue removed during the third trimester of pregnancy was  
XX screened with rabbit anti-bovine kappa-casein cDNA. The cDNA  
XX insert of a recombinant phage was amplified by PCR. The sequence  
XX of an isolated full-length clone encoding human kappa-casein  
XX (AAR72696) is given in AAQ89598.  
XX Sequence 858 BP; 266 A; 237 C; 137 G; 218 T; 0 other;  
SQ  
Query Match 78.1%; Score 16.4; DB 16; Length 858;  
Best Local Similarity 94.4%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 ggggtggaggcgtattgg 21  
DB 421 GGGTGGTGCTATTGG 404  
RESULT 15  
AAZ21101/C  
ID AAZ21101 standard; cDNA; 2185 BP.  
XX  
AC AAZ21101;  
DT 17-NOV-1999 (first entry)  
XX Human secreted protein clone qb401\_6 encoding cDNA.  
XX Human; secreted protein; biological activity; nutritional; cytokine;  
XX cell proliferation; differentiation; immune stimulating; vaccine;  
XX hematopoiesis regulation; tissue growth; haemostatic; thrombolytic;  
XX anti-inflammatory; tumour inhibition; ss.  
XX Homo sapiens.  
XX OS  
XX WO9946287-A1.  
XX  
PD 16-SEP-1999.  
XX  
XX 11-MAR-1999; 99WO-US05243.  
XX  
XX 11-MAR-1998; 98US-0077521.  
XX 14-MAY-1998; 98US-0079124.  
XX 10-MAR-1999; 99US-0266105.  
XX  
XX (GEMY ) GENETICS INST INC.  
XX  
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ;  
XX  
XX WPI; 1999-551362/46.  
XX P-PSDB; AAY29869.  
XX  
XX Polynucleotides encoding secreted human proteins, derived from human  
XX fetal brain, human adult blood, human adult bladder, or human adult

PT neural tissue cDNA libraries.  
XX Claim 26; Page 109; 118pp; English.  
XX  
CC AAZ21093 to AAZ21102 encode new human secreted proteins and AAY29861 to  
CC AAY29873 represent the secreted proteins encoded by the polynucleotide  
CC sequences. AAZ21103 to AAZ21112 represent probes for the secreted  
CC proteins. The polynucleotides and proteins are predicted to have  
CC biological activities which would make them suitable for treating,  
CC preventing or ameliorating medical conditions in humans and animals,  
CC although no supporting data is given. Suggested activities include  
CC nutritional activity, cytokine and cell proliferation/differentiation  
CC activity, immune stimulating activity, tissue growth activity,  
CC hematopoiesis regulating activity, chemotactic/chemokinetic activity, haemostatic  
CC activin/inhibin activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, and tumour  
CC inhibition activity. The polynucleotides and proteins can also be used  
CC as nutritional sources or supplements. Such uses include use as a protein  
CC or amino acid supplement, use as a carbon source, use as a nitrogen  
CC in compositions used for bone, cartilage, tendon, ligament, and/or nerve  
CC tissue growth or regeneration, as well as for wound healing and tissue  
CC repair and replacement, and in the treatment of burns, incisions and  
CC ulcers. The proteins which induce cartilage and/or bone growth in  
CC circumstances where bone is not normally formed, have application in  
CC the healing of bone fractures and cartilage damage or defects in humans  
CC and other animals.  
XX Sequence 2185 BP; 472 A; 572 C; 548 G; 593 T; 0 other;  
SQ  
Query Match 78.1%; Score 16.4; DB 20; Length 2185;  
Best Local Similarity 94.4%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 taggggtggaggcgtatttt 19  
DB 1052 TAGGGTGGAGGCTATTT 1035

Search completed: February 25, 2002, 18:17:21  
Job time: 16679 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:08 ; Search time 301.6 Seconds  
(without alignments)  
15.769 Million cell updates/sec

Title: US-09-698-903B-3  
Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTOUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21	100.0	4946	US-08-817-188-1	Sequence 1, Appli
C 2	21	100.0	5864	US-08-894-440-4	Sequence 4, Appli
C 3	16.4	78.1	857	US-08-308-883-1	Sequence 1, Appli
C 4	16.4	78.1	857	US-08-730-163-1	Sequence 1, Appli
C 5	16.4	78.1	857	US-08-256-799-1	Sequence 1, Appli
C 6	16.4	78.1	857	US-08-462-437-1	Sequence 1, Appli
C 7	16.4	78.1	858	US-07-962-569A-7	Sequence 7, Appli
C 8	16.4	78.1	9844	US-08-462-437-30	Sequence 30, Appli
C 9	16.4	78.1	13104	US-08-256-799-4	Sequence 4, Appli
C 10	16.4	78.1	13104	US-08-462-437-4	Sequence 4, Appli
C 11	16.2	77.1	1418	US-09-111-470-7	Sequence 7, Appli
C 12	16.2	77.1	24979	US-08-147-777-3	Sequence 3, Appli
C 13	16.2	77.1	24979	US-08-452-872-3	Sequence 3, Appli
C 14	16.2	77.1	24979	PCT-US93-03985-3	Sequence 3, Appli
C 15	15.8	75.2	1025	US-08-136-277-23	Sequence 23, Appli
C 16	15.8	75.2	1025	US-08-479-403-23	Sequence 23, Appli
C 17	15.8	75.2	1025	US-08-835-734-23	Sequence 23, Appli
C 18	15.8	75.2	1643	US-08-805-118-2	Sequence 2, Appli
C 19	15.8	75.2	1780	US-08-724-394A-19	Sequence 19, Appli
C 20	15.4	73.3	5769	US-08-652-971-1	Sequence 1, Appli
C 21	15.4	73.3	5769	US-08-991-258A-1	Sequence 1, Appli
C 22	15.4	73.3	5769	US-08-769-399-1	Sequence 1, Appli
C 23	15.4	73.3	5769	US-08-991-953A-1	Sequence 1, Appli
C 24	15.2	72.4	5538	US-08-231-193A-55	Sequence 55, Appli
C 25	15.2	72.4	5538	US-08-486-273A-55	Sequence 55, Appli
C 26	15.2	72.4	5538	US-08-940-086A-55	Sequence 55, Appli
C 27	15.2	72.4	9046	US-08-227-536-1	Sequence 1, Appli

C 28	15.2	72.4	9046	5	PCT-US95-04682-1	Sequence 1, Appli
C 29	14.8	70.5	354	2	US-08-676-279-38	Sequence 38, Appli
C 30	14.8	70.5	569	4	US-09-326-039-13	Sequence 13, Appli
C 31	14.8	70.5	2750	1	US-08-136-277-1	Sequence 1, Appli
C 32	14.8	70.5	2750	2	US-08-479-403-1	Sequence 1, Appli
C 33	14.8	70.5	2750	3	US-08-835-734-1	Sequence 1, Appli
C 34	14.6	69.5	35100	1	US-08-306-691B-19	Sequence 19, Appli
C 35	14.6	69.5	35100	5	PCT-US93-06251-19	Sequence 19, Appli
C 36	14.4	68.6	90050	4	US-09-245-041-5	Sequence 5, Appli
C 37	14.2	67.6	2899	2	US-08-624-581-2	Sequence 2, Appli
C 38	14.2	67.6	2959	2	US-08-624-581-1	Sequence 1, Appli
C 39	14.2	67.6	5994	3	US-09-032-365A-11	Sequence 11, Appli
C 40	14.2	67.6	13146	2	US-08-724-354D-3	Sequence 3, Appli
C 41	14.2	67.6	13146	3	US-09-270-984A-3	Sequence 3, Appli
C 42	14.2	67.6	14311	4	US-08-646-695-1	Sequence 1, Appli
C 43	14.2	67.6	14311	4	US-08-646-695-7	Sequence 7, Appli
C 44	14.2	67.6	14311	5	PCT-US96-06053-1	Sequence 1, Appli
C 45	14.2	67.6	14311	5	PCT-US96-06053-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-08-817-188-1/c  
; Sequence 1, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; CURRENT FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
; EARLIER FILING DATE: 1996-07-31  
; EARLIER APPLICATION NUMBER: EP 95401844.6  
; EARLIER FILING DATE: 1995-08-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4946  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; OTHER INFORMATION: plasmid pTHW107  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((1)..(25))  
; OTHER INFORMATION: T-DNA right border (RB)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((97)..(330))  
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
; OTHER INFORMATION: T-DNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((331)..(882))  
; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
; OTHER INFORMATION: transferase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((883)..(2608))  
; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
; OTHER INFORMATION: Promoter region of Rubisco small subunit gene of  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2658)..(3031))  
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
; OTHER INFORMATION: gene of Agrobacterium T-DNA  
; FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4876))
; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((4922)..(4946))
; OTHER INFORMATION: LB: T-DNA left border
US-08-817-188-1

Query Match      100.0%; Score 21; DB 3; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.056;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
   |||||
Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 2
US-08-894-440-4/c
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pICOLL3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Phos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4

Query Match      100.0%; Score 21; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
   |||||
Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 3
US-08-308-883-1/c
; Sequence 1, Application US/08308883
; Patent No. 5576300
; GENERAL INFORMATION:
; APPLICANT: Mukerji, P. A.
; APPLICANT: Prieto, P. A.
; APPLICANT: Seo, A. E.-Y.
; APPLICANT: Baxter, J. H.
; APPLICANT: Cummings, R.D.
; TITLE OF INVENTION: Method for Inhibition of Human Rotavirus Infection.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lonnie R. Drayer
; ADDRESSEE: ROSS Products Division
; ADDRESSEE: Abbott Laboratories
; STREET: 625 Cleveland Avenue
; CITY: Columbus
; STATE: Ohio
; COUNTRY: United States
; ZIP: 43215
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb storage
; OPERATING SYSTEM: Macintosh
; SOFTWARE: ClarisWorks 1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,883
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: No. 5576300 applicable
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (614) 624-3774
; TELEFAX: (614) 624-3074
; TELEX: No. 5576300e
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```



```

STATE: Ohio
COUNTRY: United States
ZIP: 43215
COMPUTER READABLE FORM:
MEDIA TYPE: 3.5 inch, 1.44 MB storage
OPERATING SYSTEM: Macintosh System 7
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/730.163
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US/08/308.882
FILING DATE: 16-SEP-1994
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 624-3774
TELEFAX: (614) 624-3074
TELEX: NO. 5712350e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 857 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: cDNA
DESCRIPTION: Human milk kappa-casein
HYPOTHETICAL: No
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: Human
ORGANISM: Homo sapiens
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: Adult
HAPLOTYPE:
TISSUE TYPE: Mammary gland
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: Human Mammary Gland
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:

```

```

1 NAME/KEY: CDS
2 LOCATION: 45...593
3 IDENTIFICATION METHOD: DNA sequencing and restriction analysis
4 OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO:
5 PUBLICATION INFORMATION:
6 AUTHORS: L. Hansson et al
7 TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the P
8 JOURNAL:
9 VOLUME:
10 ISSUE:
11 PAGES:
12 DATE:
13 DOCUMENT NUMBER: PCT/WO93/15196
14 FILING DATE: 25-JAN-1993
15 PUBLICATION DATE: 05-AUG-1993
16 RELEVANT RESIDUES IN SEQ ID NO:
17 US-08-730-163-1

```

Query Match 78.1%; Score 16.4; DB 1; Length 857;  
Best Local Similarity 94.4%; Pred. No. 9.5;

Qy 4 gggtgggaggctatttg 21  
|||||  
Dp 359 GGGTGGGTGGCTATTGG 342

RESULT 5  
US-08-256-799-1/C  
; Sequence 1, Application US/08256799  
; Patent No. 6222094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMQVIST, Mats  
; APPLICANT: BERGSTROM, Sven  
; APPLICANT: HERNELL, Olie  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR  
; TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/256,799  
; FILING DATE: 06-DEC-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON=1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 45..104  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 13..44  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: 594..848  
US-08-256-799-1

Query Match 78.1%; Score 16.4; DB 4; Length 857;  
Best Local Similarity 94.4%; Pred. NO. 9.5;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gggtgggaggtatttgg 21

Db 359 GGTTGGTGGCTATTGG 342  
||||| |||||||  
RESULT 6  
US-08-462-437-1/C  
; Sequence 1, Application US/08462437  
; Patent No. 6232094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMQVIST, Mats  
; APPLICANT: BERGSTROM, Sven  
; APPLICANT: HERNELL, Olie  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS  
; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,437  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON=1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 45..104  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 13..44  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: 594..848  
US-08-462-437-1

Query Match 78.1%; Score 16.4; DB 4; Length 857;  
Best Local Similarity 94.4%; Pred. NO. 9.5;

```
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 ggggtggaggctattgg 21
      ||||| |||||
Db 359 GGGTGGTGGCTATTGG 342

RESULT 7
US-07-962-569A-7/c
; Sequence 7, Application US/07962569A
; Patent No. 5391497
; GENERAL INFORMATION:
; APPLICANT: MENON, RAVI S.
; APPLICANT: JEFFERS, KATHLEEN F.
; APPLICANT: CHANG, YING-FON
; APPLICANT: HAM, RICHARD G.
; TITLE OF INVENTION: HUMAN K-CASEIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FREDERICK W. PEPPER, PH.D.
; STREET: 11545 W. BERNARDO COURT, STE. 302
; CITY: SAN DIEGO
; STATE: CA
; COUNTRY: USA
; ZIP: 92127
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/962.569A
; FILING DATE: 19921013
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PEPPER PH.D., FREDERICK W.
; REGISTRATION NUMBER: 31,286
; REFERENCE/DOCKET NUMBER: 920224.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 451-1120
; TELEFAX: (619) 451-9628
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 107..656
US-07-962-569A-7

Query Match 78.1%; Score 16.4; DB 1; Length 858;
Best Local Similarity 94.4%; Pred. No. 9.5;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 ggggtggaggctattgg 21
      ||||| |||||
Db 421 GGGTGGTGGCTATTGG 404

RESULT 8
US-08-462-437-30/c
; Sequence 30, Application US/08462437
; Patent No. 6232094
; GENERAL INFORMATION:
; APPLICANT: HANSSON, Lennart
; APPLICANT: STROEMQVIST, Mats
; APPLICANT: BERGSTROEM, Sven
; APPLICANT: HERNELL, Olle
; APPLICANT: TOERNELL, Jan
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS
```

```

; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462.437
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 88/92
; FILING DATE: 23-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: HANSSON-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9844 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 362..425
; FEATURE:
; NAME/KEY: intron
; LOCATION: 426..2571
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2572..2633
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2634..5575
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5576..5608
; FEATURE:
; NAME/KEY: intron
; LOCATION: 5609..5754
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5755..7250
; FEATURE:
; NAME/KEY: intron
; LOCATION: 7251..9017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 9018..9184
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2580..2633, 5576..5608, 6755..7216)
US-08-462-437-30

Query Match 78.1%; Score 16.4; DB 4; Length 9844;
```



; LOCATION: 10511..12277  
; FEATURE: exon  
; NAME/KEY: exon  
; LOCATION: 8835..8867  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 10015..10510  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 12278..12443  
US-08-462-437-4

Query Match 78.1%; Score 16.4; DB 4; Length 13104;  
Best Local Similarity 94.4%; Pred. No. 14;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggggtggaggctatttgg 21  
||||| |||||||  
Db 10242 GGGTGGTGCTATTGG 10225

RESULT 11  
US-09-111-470-7  
; Sequence 7, Application US/09111470  
; Patent No. 6277959  
; GENERAL INFORMATION:  
; APPLICANT: Valladeau, Jenny  
; APPLICANT: Ravel, Odile  
; APPLICANT: Bates, Elizabeth E.M.  
; APPLICANT: Ford, John  
; APPLICANT: Saeland, Sem  
; APPLICANT: Lebecque, Serge J.E.  
; TITLE OF INVENTION: Mammalian Membrane Protein Genes;  
; TITLE OF INVENTION: Related Reagents  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/111.470  
; FILING DATE: 08-JUL-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/053,080  
; FILING DATE: 09-JUL-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: SF0695  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650)852-9196  
; TELEFAX: (650)496-1200  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1418 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 279..992  
; FEATURE:

; NAME/KEY: misc\_feature  
; LOCATION: 1348  
; OTHER INFORMATION: /note= "poly-A addition motif"  
US-09-111-470-7

Query Match 77.1%; Score 16.2; DB 4; Length 1418;  
Best Local Similarity 85.7%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atagggtggaggctatttgg 21  
||||| |||||||  
Db 602 AAAGATTGGAGGCTATTGG 622

RESULT 12  
US-08-147-777-3  
; Sequence 3, Application US/08147777  
; Patent No. 5914265  
; GENERAL INFORMATION:  
; APPLICANT: Roop, Dennis R.  
; APPLICANT: Rothnagel, Joseph A.  
; APPLICANT: Greenhalgh, David A.  
; APPLICANT: Yuspa, Stuart H.  
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LYON & LYON  
; STREET: 611 West Sixth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90017  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/147,777  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: described below: two  
; APPLICATION NUMBER: 07/876,289  
; FILING DATE: April 30, 1992  
; APPLICATION NUMBER: Unassigned (204/144)  
; FILING DATE: October 29, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 204/153  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 24979 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-147-777-3

Query Match 77.1%; Score 16.2; DB 2; Length 24979;  
Best Local Similarity 85.7%; Pred. No. 19;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atagggtggaggctatttgg 21

```
Db 3248 AAAGGTGGGAAGCTATATGG 3268
|||||
RESULT 13
US-08-452-872-3
; Sequence 3, Application US/08452872
; Patent No. 6057298
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,872
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,777
; FILING DATE:
; APPLICATION NUMBER: 07/876,289
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: Unassigned (204/144)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/153
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24979 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-452-872-3

Query Match 77.1%; Score 16.2; DB 3; Length 24979;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
| |||||
Db 3248 AAAGGTGGGAAGCTATATGG 3268

RESULT 14
US-08-136-277-23
; Sequence 23, Application US/08136277
; Patent No. 5644045
; GENERAL INFORMATION:
; APPLICANT: MANDEL, Jean-Louis
; APPLICANT: AUBOURG, Patrick
; APPLICANT: MOSSER, Jean
; APPLICANT: SARDE, Claude
; TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
; TITLE OF INVENTION: CORRESPONDING PROTEIN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

PCT-US93-03985-3

Query Match 77.1%; Score 16.2; DB 5; Length 24979;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
| |||||
Db 3248 AAAGGTGGGAAGCTATATGG 3268

RESULT 15
US-08-136-277-23
; Sequence 23, Application US/08136277
; Patent No. 5644045
; GENERAL INFORMATION:
; APPLICANT: MANDEL, Jean-Louis
; APPLICANT: AUBOURG, Patrick
; APPLICANT: MOSSER, Jean
; APPLICANT: SARDE, Claude
; TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
; TITLE OF INVENTION: CORRESPONDING PROTEIN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

PCT-US93-03985-3
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,277
; FILING DATE: 15-OCT-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: B2272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1025 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-136-277-23

```

```

Query Match      75.2%; Score 15.8; DB 1; Length 1025;
Best Local Similarity 89.5%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 taggggtggaggctatttg 20
        |||||
Db      148 TAGGTTGGGAGGCTATGTG 166

```

Search completed: February 25, 2002, 18:05:11  
Job time: 18559 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:20:51 ; Search time 8261.74 Seconds  
(without alignments)  
27.314 Million cell updates/sec

Title: US-09-698-903b-3

Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estfun:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estom:\*
  - 5: em\_estpl:\*
  - 6: em\_estba:\*
  - 7: em\_estro:\*
  - 8: em\_estov:\*
  - 9: em\_htc:\*
  - 10: gb\_est1:\*
  - 11: gb\_est2:\*
  - 12: gb\_hc:\*
  - 13: gb\_gss:\*
  - 14: em\_gss\_fun:\*
  - 15: em\_gss\_hum:\*
  - 16: em\_gss\_inv:\*
  - 17: em\_gss\_pln:\*
  - 18: em\_gss\_pro:\*
  - 19: em\_gss\_rod:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	Description
c 1	17.4	82.9	212 11 BG983367
c 2	17.4	82.9	429 11 BF924973
c 3	17.4	81.0	501 10 AW638905
c 4	17.4	81.0	583 10 AW635936
c 5	16.8	80.0	285 10 AV145846
c 6	16.8	80.0	321 11 F05693
c 7	16.8	80.0	424 11 W57725
c 8	16.8	80.0	429 11 B1188834
c 9	16.8	80.0	587 11 BG494856
c 10	16.8	80.0	626 10 AW644184
c 11	16.8	80.0	667 10 AW639842
c 12	16.8	80.0	892 11 BF167952

c 13	16.8	80.0	914 13 CNS03N93
c 14	16.4	78.1	184 13 AZ016369
c 15	16.4	78.1	284 11 BG210286
c 16	16.4	78.1	385 10 AW736851
c 17	16.4	78.1	462 11 R41615
c 18	16.4	78.1	575 10 AW578936
c 19	16.4	78.1	658 13 BH036913
c 20	16.4	78.1	714 11 BG217189
c 21	16.4	78.1	728 11 BG204492
c 22	16.4	78.1	781 13 AZ188345
c 23	16.4	78.1	798 11 EG199886
c 24	16.4	78.1	822 11 EG194262
c 25	16.4	78.1	829 11 EG199887
c 26	16.4	78.1	837 11 BG211849
c 27	16.4	78.1	838 11 BG209779
c 28	16.4	78.1	850 11 BG215054
c 29	16.4	78.1	859 11 BG219740
c 30	16.4	78.1	866 11 BG205014
c 31	16.4	78.1	962 11 EG195797
c 32	16.2	77.1	162 10 BE011473
c 33	16.2	77.1	166 11 BF328785
c 34	16.2	77.1	228 10 AV338030
c 35	16.2	77.1	263 10 AA178060
c 36	16.2	77.1	265 10 BB252853
c 37	16.2	77.1	270 10 BB307020
c 38	16.2	77.1	274 10 BB008234
c 39	16.2	77.1	298 10 BB464468
c 40	16.2	77.1	321 10 AA850434
c 41	16.2	77.1	323 10 BB319109
c 42	16.2	77.1	326 10 BB319885
c 43	16.2	77.1	326 10 BB461095
c 44	16.2	77.1	348 13 AQ610088
c 45	16.2	77.1	385 10 BE113833

ALIGNMENTS

RESULT 1	1	212 bp	mrna	EST	12-JUN-2001
LOCUS	BG983367/c				
DEFINITION	IL5-CN0067-060301-384-fl2 CN0067 Homo sapiens cdna, mRNA sequence.				
ACCESSION	BG983367				
VERSION	BG983367.1	GI:14386102			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 212) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunschein,A., deOliveira,P., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-CN0067-060301-fl2&t3=2001-03-06&t4=1) Seq primer: puc 18 forward				

High quality sequence stop: 211.

FEATURES  
source  
1. .212  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NC0067"  
/dev\_stage="Adult"  
/note="Organ: colon\_normal; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
38 a 84 c 32 g 58 t

BASE COUNT  
ORIGIN

Query Match 82.9%; Score 17.4; DB 11; Length 212;  
Best Local Similarity 94.7%; Pred. No. 6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 2 taggggtggaggctatttg 20  
|||||  
Db 108 TAGGGTGGGAGGCTATGTG 90

RESULT 2

LOCUS BF924973 429 bp mRNA EST 19-JAN-2001  
DEFINITION IL2-NT0200-281100-263-C11 NT0200 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF924973  
VERSION BF924973.1 GI:12320948  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 429)  
Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&t2=IL2-NT0200-  
281100-263-C11&t3=2000-11-28&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 351.

FEATURES  
source

1. .429  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NT0200"  
/dev\_stage="Adult"  
/note="Organ: nervous\_tumor; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
99 a 106 c 132 g 90 t 2 others

BASE COUNT  
ORIGIN

Query Match 82.9%; Score 17.4; DB 11; Length 429;  
Best Local Similarity 94.7%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Caps 0;  
Qy 3 aggggtggaggctatttg 21  
|||||  
Db 392 AGCGTGGGAGGCTATTTGG 410

RESULT 3

LOCUS AW638905/c 501 bp mRNA EST 26-APR-2001  
DEFINITION bl76B06.w1 Blackshear/Soares normalized Xenopus egg library Xenopus  
laevis cDNA clone PBX0076B06 5', mRNA sequence.  
ACCESSION AW638905  
VERSION AW638905.1 GI:7396075  
KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus.  
1 (bases 1 to 501)  
REFERENCE 1  
AUTHORS Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.,  
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman  
J.W., Bonaldo,M.F. and Soares,M.B.  
The NIH's Xenopus maternal EST project: interim analysis of the  
first 13,879 ESTs from unfertilized eggs  
Gene 267 (1), 71-87 (2001)  
21211403  
Contact: Perry J. Blackshear  
Office of Clinical Research and Laboratory of Signal Transduction  
National Institute of Environmental Health Sciences  
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,  
USA  
Tel: 919 541-4899  
Fax: 919 541-4571  
Email: black009@niehs.nih.gov  
Clone is available through Research Genetics, Inc., 2130 Memorial  
Parkway, Huntsville, AL 35901  
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email  
cdna@resgen.com  
DNA Sequencing and analyses performed by National Institutes of  
Health Intramural Sequencing Center (NISC).  
PCR PRIMERS  
FORWARD: TGTAACGACGCGCCAGT  
BACKWARD: CAGGAACAGCTATGACC  
Plate: 0076 row: B column: 06  
Seq primer: T7 primer.

FEATURES  
source

1. 501  
Location/Qualifiers  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="PBX0076B06"  
/clone\_lib="Blackshear/Soares normalized Xenopus egg  
library"  
/sex="female"  
/tissue\_type="unfertilized egg"  
/cell\_type="unfertilized egg"  
/dev\_stage="unfertilized egg"  
/lab\_host="DH10B"  
/note="Vector: pT7T3-Pac; Site\_1: EcoRI; Site\_2: NotI;  
PolyA-selected mRNA was prepared from unfertilized Xenopus  
laevis eggs. The library was constructed in the vector  
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and  
Soares, M.B. Normalization and subtraction: two

approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dri18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7T3-Pac vector. The library contained approximately 7.2 X 10<sup>5</sup> recombinants, with average insert sizes of 1-1.5 kb."

```

BASE COUNT      161 a  116 c  106 g  118 t
ORIGIN

Query Match      81.0%; Score 17; DB 10; Length 501;
Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 taggggtggaggctatt 18
|||||
Db 317 TAGGGTGGGAGGCTATT 301
|||||

RESULT 4
AW635936/c
LOCUS      583 bp      mRNA      EST      26-APR-2001
DEFINITION bla04003.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
laevis cDNA clone PBX0040D03 5', mRNA sequence.
ACCESSION  AW635936
VERSION     AW635936.1 GI:7393017
KEYWORDS   EST,
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 583)
Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
,J.W., Bonaldo,M.F. and Soares,M.B.
The NIEHS Xenopus maternal EST project: Interim analysis of the
first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
21211403
Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com
DNA sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAACGACGCGCCAGT
BACKWARD: CAGGAACAGCTATGACC
Plate: 0040 row: D column: 03
Seq primer: T7 primer.
Location/Qualifiers
1..583
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0040D03"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;

FEATURES
source
1..285
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2810449H10"
/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"
63 a 87 c 53 g 82 t
BASE COUNT
ORIGIN

```

PolyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dri18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7T3-Pac vector. The library contained approximately 7.2 X 10<sup>5</sup> recombinants, with average insert sizes of 1-1.5 kb."

```

BASE COUNT      185 a  126 c  128 g  144 t
ORIGIN

Query Match      81.0%; Score 17; DB 10; Length 583;
Best Local Similarity 100.0%; Pred. No. 9.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 taggggtggaggctatt 18
|||||
Db 433 TAGGGTGGGAGGCTATT 417
|||||

RESULT 5
AV145846/c
LOCUS      285 bp      mRNA      EST      03-JUL-1999
DEFINITION AV145846 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
clone 2810449H10, mRNA sequence.
ACCESSION  AV145846
VERSION     AV145846.1 GI:5349979
KEYWORDS   EST,
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 285)
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H., Owa,C.,
Sato,K., Shibata,Y., Shigenoto,Y., Shiraki,T., Sogabe,Y., Sugahara
,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomimaga,N.,
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermolabile and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..285
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2810449H10"
/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"
63 a 87 c 53 g 82 t
BASE COUNT
ORIGIN

```

```

Query Match      80.0%; Score 16.8; DB 10; Length 285;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggggtggaggctattgg 21
|||||
Db 47 TAGGGTGGGAGCAATTAGG 28

RESULT 6
LOCUS F05693/c
DEFINITION HSC0JC091 normalized infant brain cDNA Homo sapiens cDNA clone
C-0jc09, mRNA sequence.
ACCESSION F05693
VERSION F05693.1 GI:669509
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 321)
AUTHORS Aufrey,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
,M.D., Duprat,S., Houligatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_id: C; Genexpress_sequence_id: ylc-0jc09
Seq primer: (-21)M13_universal.
FEATURES
source
1..321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="c-0jc09"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/notes="Organ: brain; Vector: lafm1d BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate=muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafm1d BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press."
BASE COUNT 108 a 53 c 46 g 114 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 11; Length 321;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggggtggaggctattgg 21
|||||
Db 106 TGGGGTGGGAGCAATTGG 87

RESULT 7
LOCUS W57725/c
DEFINITION zd20c10.sl Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:341202 3', mRNA sequence.
ACCESSION W57725
VERSION W57725.1 GI:1364441
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 424)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 866 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 376.
FEATURES
source
1..424
/organism="Homo sapiens"
/db_xref="GDB:1266577"
/db_xref="taxon:9606"
/clone_lib="IMAGE:341202"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGGAGCGGCATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."
BASE COUNT 145 a 87 c 77 g 115 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 11; Length 424;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 20
|||||
Db 374 ATGGGTGGGAGGCTCTTG 355

RESULT 8
LOCUS BI188834/c
DEFINITION BI188834 429 bp mRNA
EST 10-JUL-2001
d3d12fs.fl Fusarium sporotrichioides Tri 10 overexpressed cDNA
library Fusarium sporotrichioides cDNA clone d3d12fs 3', mRNA
sequence.
ACCESSION BI188834
VERSION BI188834.1 GI:14662513
KEYWORDS EST.
SOURCE Fusarium sporotrichioides.

```

ORGANISM Fusarium sporotrichioides  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreales; mitosporic Hypocreales; Fusarium.

REFERENCE 1 (bases 1 to 429)  
AUTHORS Ren, Q., Tag, A., Peplow, A., Lai, H., Kupfer, C., Peterson, A., Beremand, M. and Roe, B.  
TITLE Analysis of a Fusarium sporotrichioides EST database  
JOURNAL Unpublished (2001)  
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
Department of Chemistry and Biochemistry  
Advanced Center for Genome Technology, University of Oklahoma  
620 Parrington Oval, Norman, OK 73019, USA  
Tel: 405 325 4912  
Fax: 405 325 7762  
Email: broe@ou.edu  
Contact Dr. Marian Beremand regarding clone availability. Included  
is the best homolog from a blastx search of Genbank nr 04-09-01  
71 3 9 g1|3550690|emb|CAA09 (AJ010199) variable surface  
glycoprotein|T  
Seq primer: M13-20  
High quality sequence stop: 405.  
Location/Qualifiers  
1..429  
/organism="Fusarium sporotrichioides"  
/strain="Tri 10"  
/db\_xref="taxon:5514"  
/clones="d3d12fs"  
/clone\_lib="Fusarium sporotrichioides Tri 10 overexpressed  
cDNA library"  
/note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript  
; 3' end of cDNA cloned into XhoI site of pBluescript"  
BASE COUNT 137 a 102 c 68 g 122 t  
ORIGIN  
1 atagggtggaggctatttg 20  
|||||  
Db 164 ATTGGAGGAGGCTATTG 145

RESULT 9  
LOCUS BG494856 587 bp mRNA EST 27-MAR-2001  
DEFINITION 602540862f1 NIH\_MGC\_59 Homo sapiens cDNA clone IMAGE:4671903 5',  
mRNA sequence.  
ACCESSION BG494856  
VERSION BG494856.1 GI:13456371  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 587)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaaps-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: CLONETECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM1487 row: 1 column: 16  
High quality sequence stop: 584.  
Location/Qualifiers

FEATURES  
source

1..587  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clones="IMAGE:4671903"  
/clone\_lib="NIH\_MGC\_59"  
/tissue\_type="mucoepidermoid carcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggccgctcgcc); Site\_2: SfiI (ggccattggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CAGGCCATTATGGC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGGAGCGGCACATG-dT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
Library."  
BASE COUNT 197 a 104 c 87 g 199 t  
ORIGIN  
1 taggggtggaggctatttg 21  
|||||  
Db 191 TGGGTGGGAGCAATTGG 172

RESULT 10  
LOCUS AW644184 626 bp mRNA EST 26-APR-2001  
DEFINITION cm38d02.w1 Blackshear/Soares normalized Xenopus egg library Xenopus  
laevis cDNA clone PBX0137D02 5', mRNA sequence.  
ACCESSION AW644184  
VERSION AW644184.1 GI:7401561  
KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 626)  
AUTHORS Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.,  
Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman,  
J.W., Bonaldo, M.F. and Soares, M.B.  
TITLE The NIHXS Xenopus maternal EST project: interim analysis of the  
first 13,879 ESTs from unfertilized eggs  
JOURNAL Gene 267 (1), 71-87 (2001)  
MEDLINE 21211403  
COMMENT Contact: Perry J. Blackshear  
Office of Clinical Research and Laboratory of Signal Transduction  
National Institute of Environmental Health Sciences  
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,  
USA  
Tel: 919 541-4899  
Fax: 919 541-4571  
Email: black009@niehs.nih.gov  
Clone is available through Research Genetics, Inc., 2130 Memorial  
Parkway, Huntsville, AL 35901  
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email  
cdna@resgen.com  
DNA Sequencing and analyses performed by National Institutes of  
Health Intramural Sequencing Center (NISC).  
PCR Primers  
FORWARD: TGTAACGACGCGCCAGT  
BACKWARD: CAGGAACAGCTATGACC  
Plate: 0137 row: D column: 02  
Seq primer: T7 primer.  
Location/Qualifiers

FEATURES  
source

source

```

1. .626
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0137D02"
/library="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      204 a 116 c 131 g 175 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 10; Length 626;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggtggaggctattgg 21
|||||
Db 131 TAGGGTGGGAGGATATTAGG 150

RESULT 11
AW639842
LOCUS      AW639842      667 bp      mRNA      EST      26-APR-2001
DEFINITION b187h11.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
laevis cDNA clone PBX0087H11 5', mRNA sequence.
ACCESSION      AW639842
VERSION
KEYWORDS
SOURCE
ORGANISM      African clawed frog.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 667)
Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
J.W., Bonaldo,M.F. and Soares,M.B.
the NIEHS Xenopus maternal EST project: interim analysis of the
first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
21211403
Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
42-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR PRIMERS
FORWARD: TGTAACACGCGCCAGT

```

```

BACKWARD: CAGGAACAGCTATGACC
Plate: 0087 row: H column: 11
Seq primer: T7 primer
FEATURES
source
1. .667
Location/Qualifiers
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0087H11"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      214 a 124 c 144 g 185 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 10; Length 667;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggtggaggctattgg 21
|||||
Db 143 TAGGGTGGGAGGATATTAGG 162

RESULT 12
BF167952/c
LOCUS      BF167952      892 bp      mRNA      EST      30-OCT-2000
DEFINITION 601775688F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:4017644 5',
mRNA sequence.
ACCESSION      BF167952
VERSION
KEYWORDS
SOURCE
ORGANISM      house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 892)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9266 row: o column: 21
High quality sequence stop: 662.
Location/Qualifiers
1. .892
/organism="Mus musculus"
/strain="Czech II (fetal)"
/db_xref="taxon:10090"
/clone="IMAGE:4017644"
FEATURES
source
1. .892
Location/Qualifiers

```

```

/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site.1: Sali;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT      241 a   225 c   237 g   189 t
ORIGIN

Query Match      80.08; Score 16.8; DB 11; Length 892;
Best Local Similarity 90.08; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2 tagggtagggagctattgg 21
||||| ||||||| |||
Db  871 TAGGGTGTGAGGCATATGG 852

RESULT 13
CNS03N93/c
LOCUS      CNS03N93      914 bp      DNA      GSS      17-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
039K16 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION   AL251760.1 GI:797272
VERSION     AL251760.1 GI:797272
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 914)
AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 914)
AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE      Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 914)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL     Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source           1..914
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="039K16"
                     /clone_lib="G"
BASE COUNT      202 a   257 c   239 g   212 t   4 others
ORIGIN

Query Match      80.08; Score 16.8; DB 13; Length 914;
Best Local Similarity 85.78; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy  3 agggtagggagcctattg 20
||||| ||||||| |||
Db  60 AGGGTGGGAGGCAATTG 77

RESULT 15
BG210286/c
LOCUS      BG210286      284 bp      mRNA      EST      21-APR-2001
DEFINITION RST29821 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.

```

```

Qy  1 atagggtggggagctattgg 21
||||| ||||||| |||
Db  281 ATGGGGTGGGAGGCNACTTGG 261

RESULT 14
AZ016369
LOCUS      AZ016369      184 bp      DNA      GSS      25-FEB-2000
DEFINITION RPCI-23-264M11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-264M11
, DNA sequence.
ACCESSION   AZ016369
VERSION     AZ016369.1 GI:7091753
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 184)
AUTHORS    Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsagaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL     Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-264M11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 264 row: M column: 11
Seq primer: Sp6
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..184
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-264M11"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site.1:
EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      40 a   36 c   49 g   59 t
ORIGIN

Query Match      78.18; Score 16.4; DB 13; Length 184;
Best Local Similarity 94.48; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  3 agggtagggagcctattg 20
||||| ||||||| |||
Db  60 AGGGTGGGAGGCAATTG 77

RESULT 15
BG210286/c
LOCUS      BG210286      284 bp      mRNA      EST      21-APR-2001
DEFINITION RST29821 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.

```

ACCESSION BG210286  
VERSION BG210286.1 GI:13731973  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 284)  
Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,  
Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,  
Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Mays, R., Smith, E.,  
Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,  
and Ducar, M.  
TITLE Creation of Genome-wide Protein Expression Libraries using Random  
JOURNAL Activation of Gene Expression  
COMMENT Nat. Biotechnol. 19 (5), 440 (2001) In press  
Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scaine@atersys.com  
High quality sequence stop: 234.  
FEATURES  
Source  
1..284  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Athersys RAGE Library"  
/cell\_line="HT1080"  
/note="See 'Creation of Genome-wide Protein Expression  
Libraries using Random Activation of Gene Expression',  
Nature Biotechnology, in press. Note that even though the  
cell type indicated is HT1080, since a random activation  
method was used, these sequence tags are not necessarily  
expressed in HT1080 under normal circumstances."  
BASE COUNT 88 a 82 c 43 g 71 t  
ORIGIN  
Query Match 78.1%; Score 16.4; DB 11; Length 284;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 ggggtggaggctatttgg 21  
||||| |||||||  
Db 238 GGGTGGGTGGCTATTGG 221  
Search completed: February 25, 2002, 17:20:54  
Job time: 16147 sec





```

misc_feature 1
/note="n is a, t, g, or c"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGT 15

RESULT 2
AX063588 AX063588 15 bp DNA PAT 24-JAN-2001
LOCUS Sequence 7 from Patent WO0100801.
DEFINITION AX063588
ACCESSION AX063588
VERSION AX063588.1 GI:12541312
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 15)
AUTHORS Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 7 04-JAN-2001;
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic Oligonucleotide"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGT 15

RESULT 3
AX127751 AX127751 15 bp DNA PAT 15-MAY-2001
LOCUS Sequence 4 from Patent WO0131042.
DEFINITION AX127751
ACCESSION AX127751
VERSION AX127751.1 GI:14134398
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 15)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 4 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer MDB285"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGT 15

RESULT 4
AX167666 AX167666 15 bp DNA PAT 03-JUL-2001
LOCUS Sequence 11 from Patent WO0144277.
DEFINITION AX167666
ACCESSION AX167666
VERSION AX167666.1 GI:14597053
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 15)
AUTHORS Wegrich Glover,L., Budziszewski,G.J., Levin,J.Z. and Zhou,Q.
TITLE Herbicide target genes and methods
JOURNAL Patent: WO 0144277-A 11 21-JUN-2001;
Syngenta Participations AG (CH)
FEATURES
source Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Oligonucleotide"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGT 15

RESULT 5
E08065 E08065 15 bp DNA PAT 29-SEP-1997
LOCUS Non-specific primer.
DEFINITION E08065
ACCESSION E08065
VERSION E08065.1 GI:2176190
KEYWORDS JP 1994253843-A/4.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 15)
AUTHORS Riyuu,Y., Robaato,E.U. and Mitsukawa,N.
TITLE METHOD FOR AMPLIFYING DNA SEQUENCE
JOURNAL Patent: JP 1994253843-A 4 13-SEP-1994;
CHIKUO KANKYO SANGYO GIJUTSU KENKYU KIKO, MITSUI GIYOUSAI
SHOKUBUTSU BIO KENKYUSHO:KK
COMMENT
OS None
OC Artificial sequences.
PN JP 1994253843-A/4
PD 13-SEP-1994
PF 03-MAR-1993 JP 1993043006
PI RIYUU YOKUO, ROBAATO EFU UITSUTEIA, MITSUKAWA NORIHIRO PC
C12N15/10,C12Q1/68;
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FH source 1..15
FT /organism="Artificial sequences".

```

```

FEATURES
  source      Location/Qualifiers
    1..15
    /organism="unidentified"
    /db_xref="taxon:32644"
  BASE COUNT      1 a      1 c      3 g      5 t      5 others
  ORIGIN

  Query Match      82.7%; Score 12.4; DB 6; Length 15;
  Best Local Similarity 100.0%; Pred. No. 2e+03;
  Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  Qy 2 tcgastwtsgwgtt 15
  Db 2 TCGASTWTSGWGT 15

  RESULT 6
  MMU80260
  LOCUS      MMU80260      206 bp      DNA      ROD      19-DEC-1996
  DEFINITION Mus musculus clone ma9 L1 retrotransposon LINE1 repeat region.
  ACCESSION  U80260
  VERSION     U80260.1 GI:1737239
  KEYWORDS    house mouse.
  SOURCE      Mus musculus
  ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
  REFERENCE  1 (bases 1 to 206)
  AUTHORS    Woodcock,D.M. and Linzenmeyer,M.E.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-NOV-1996) Research, Peter MacCallum Cancer Institute,
              St. Andrew's Place, East Melbourne, Vic 3000, Australia
  FEATURES
    source      Location/Qualifiers
      1..206
      /organism="Mus musculus"
      /transposon="L1 retrotransposon"
      /db_xref="taxon:10090"
      /clone="ma9"
    repeat_region <1..>206
    BASE COUNT      32 a      14 c      74 g      86 t
    ORIGIN

  Query Match      82.7%; Score 12.4; DB 10; Length 206;
  Best Local Similarity 71.4%; Pred. No. 2e+03;
  Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

  Qy 2 tcgastwtsgwgtt 15
  Db 45 TCGAGTTTCGAGTT 58

  RESULT 7
  GMU26701/c
  LOCUS      GMU26701/c      240 bp      DNA      PLN      26-JUL-2001
  DEFINITION Glycine max satellite STR120-B.1.
  ACCESSION  U26701
  VERSION     U26701.1 GI:1142703
  KEYWORDS    soybean.
  SOURCE      Glycine max
  ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eutrosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
              Glycine.
  REFERENCE  1 (bases 1 to 240)
  AUTHORS    Morgante,M., Jurman,I., Shi,L., Zhu,T., Keim,P. and Rafalski,J.A.
  TITLE      The STR120 satellite DNA of soybean: organization, evolution and
              chromosomal specificity
  JOURNAL    Chromosome Res. 5 (6), 363-373 (1997)
  MEDLINE    98030220

```

```

PUBMED 9364938
REFERENCE 2 (bases 1 to 240)
AUTHORS  Morgante,M. and Rafalski,J.A.
TITLE    Direct Submission
JOURNAL  Submitted (09-MAY-1995) Michele Morgante, Dip. di Produzione
          Vegetale, Universita di Udine I-33100, Via delle scienze, 208,
          Udine I-33100, Italy
FEATURES
  source      Location/Qualifiers
    1..240
    /organism="Glycine max"
    /cultivar="Ripley"
    /db_xref="taxon:3847"
  repeat_unit 1..120
  satellite    1..240
  repeat_unit 121..240
  BASE COUNT  81 a      49 c      60 g      50 t
  ORIGIN

  Query Match      82.7%; Score 12.4; DB 8; Length 240;
  Best Local Similarity 71.4%; Pred. No. 2e+03;
  Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

  Qy 2 tcgastwtsgwgtt 15
  Db 132 TCGAGTTTCGAGTT 119

  RESULT 8
  PUMPAL2A
  LOCUS      PUMPAL2A      411 bp      DNA      PLN      21-JUL-1995
  DEFINITION Petroselinum crispum phenylalanine ammonia-lyase (PAL-2) gene,
              promoter region and 5' end of cds.
  ACCESSION  L37355
  VERSION     L37355.1 GI:567856
  KEYWORDS    phenylalanine ammonia-lyase.
  SOURCE      Petroselinum crispum DNA.
  ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum.
  REFERENCE  1 (bases 1 to 411)
  AUTHORS    Logemann,E., Parniske,M. and Hahlbrock,K.
  TITLE      Modes of expression and common structural features of the complete
              phenylalanine ammonia-lyase gene family in parsley
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 92 (13), 5905-5909 (1995)
  MEDLINE    95320184
  FEATURES
    source      Location/Qualifiers
      1..411
      /organism="Petroselinum crispum"
      /db_xref="taxon:4043"
      /note="(vector lambda EMBL4)"
    promoter    1..260
    TATA_signal 230..235
    gene        409..411
    /gene="PAL-2"
  BASE COUNT  128 a      99 c      40 g      144 t
  ORIGIN

  Query Match      82.7%; Score 12.4; DB 8; Length 411;
  Best Local Similarity 71.4%; Pred. No. 2e+03;
  Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

  Qy 2 tcgastwtsgwgtt 15
  Db 363 TCGAGTTTCGAGTT 376

  RESULT 9
  AX127755
  LOCUS      AX127755      415 bp      DNA      PAT      15-MAY-2001

```

DEFINITION Sequence 8 from Patent WO0131042.

ACCESSION AX127755

VERSION AX127755.1 GI:14134402

KEYWORDS synthetic construct.

SOURCE synthetic construct.

ORGANISM artificial sequence.

REFERENCE 1 (bases 1 to 415)

AUTHORS Weston,B. and de Beuckeleer,M.

TITLE Male-sterile brassica plants and methods for producing same

JOURNAL Patent: WO 0131042-A 8 03-MAY-2001;

Aventis CropScience N.V. (BE)

FEATURES

Location/Qualifiers

1..415

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="5' border flanking region of elite event MS-B2"

misc\_feature

1..234

/note="plant DNA"

misc\_feature

235..415

/note="T-DNA"

BASE COUNT 154 a 55 c 70 g 136 t

ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 415;

Best Local Similarity 71.4%; Pred. No. 2e+03;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 2 TCGAGTTGGTGTT 15

||||:|:|:|

RESULT 10

HSMCRP1/c

LOCUS Homo sapiens MHC class I-related protein MRL precursor (MRL) gene,

DEFINITION HSMCRP1 636 bp DNA PRI 11-NOV-1998

signal peptide.

ACCESSION AF073484

VERSION AF073484.1 GI:3859892

KEYWORDS 1 of 2

SEGMENT human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Yamaguchi,H., Kurosawa,Y. and Hashimoto,K.

TITLE I-related genes, human MRL and its murine ortholog

JOURNAL Biochem. Biophys. Res. Commun. 250 (3), 558-564 (1998)

MEDLINE 99003494

REFERENCE 2 (bases 1 to 636)

AUTHORS Hashimoto,K.

TITLE Direct Submission

JOURNAL Submitted (23-JUN-1998) Institute for Comprehensive Medical

Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan

FEATURES

Location/Qualifiers

1..636

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/tissue\_type="placenta"

/chromosome="1"

/map="1q25.3"

<1..45

/gene="MRL"

BASE COUNT 160 a 140 c 138 g 198 t

ORIGIN

sig\_peptide

<1..45

Query Match 82.7%; Score 12.4; DB 9; Length 636;

Best Local Similarity 71.4%; Pred. No. 2e+03;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 582 TCGAGTTTCGTGTT 595

||||:|:|:|

RESULT 12

GRO251757/c

LOCUS Globodera rostochiensis mRNA for for hypothetical protein (clone

DEFINITION AJ251757

ACCESSION AJ251757

VERSION AJ251757.1 GI:12227170

KEYWORDS ORF.

SOURCE Globodera rostochiensis.

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 88 TCGACTTTGGAGTT 75

||||:|:|:|

RESULT 11

AF283269

LOCUS Anopheles gambiae ribosomal protein S26 (Irps26) mRNA, complete

DEFINITION AF283269 676 bp mRNA INV 16-OCT-2000

cds.

ACCESSION AF283269

VERSION AF283269.1 GI:10242303

KEYWORDS African malaria mosquito.

SOURCE Anopheles gambiae

ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

REFERENCE Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;

Culicoidea; Anopheles.

1 (bases 1 to 676)

AUTHORS Oduol,F., Xu,J., Niare,O., Natarajan,R. and Vernick,K.D.

TITLE Genes identified by an expression screen of the vector mosquito

JOURNAL malaria parasites and bacteria

PUBMED Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11397-11402 (2000)

REFERENCE 11005829

2 (bases 1 to 676)

AUTHORS Oduol,F.O., Xu,J., Niare,O., Natarajan,R. and Vernick,K.D.

TITLE Direct Submission

JOURNAL Submitted (28-JUN-2000) Department of Medical and Molecular

Parasitology, New York University School of Medicine, 341 East 25th

Street, New York, NY 10010, USA

FEATURES

Location/Qualifiers

1..676

/organism="Anopheles gambiae"

/strain="G3"

/db\_xref="taxon:7165"

/dev\_stage="adult"

1..676

/gene="Irps26"

30..374

/gene="Irps26"

/note="immune-responsive"

/codon\_start=1

/product="ribosomal protein S26"

/protein\_id="AAG15374.1"

/db\_xref="GI:10242304"

/translation="MLERRNGRCCKHNRGHVKAIVCTNCARCPKDKAIKKFVIRNIV

EAAAVRDIDSDASVYSSVYLPKLYAKLHVCVAISHKVVNRNRSETRIRTPQSRFP

KDMNRQQAQRK"

BASE COUNT 160 a 179 c 195 g 142 t

ORIGIN

Query Match 82.7%; Score 12.4; DB 3; Length 676;

Best Local Similarity 71.4%; Pred. No. 2e+03;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 582 TCGAGTTTCGTGTT 595

||||:|:|:|

RESULT 12

GRO251757/c

LOCUS Globodera rostochiensis mRNA for for hypothetical protein (clone

DEFINITION AJ251757

ACCESSION AJ251757

VERSION AJ251757.1 GI:12227170

KEYWORDS ORF.

SOURCE Globodera rostochiensis.



BASE COUNT 273 a 173 c 176 g 305 t  
ORIGIN

## Query Match

Best Local Similarity 82.7%; Score 12.4; DB 8; Length 927;  
Matches 10; Conservative 71.4%; Pred. No. 2e+03;

Mismatches 4; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 694 TCGAGTTTCGAGTT 681

## RESULT 15

AF328994/c

LOCUS AF328994 942 bp DNA PLN 17-MAY-2001

DEFINITION Arabidopsis lyrata putative S-glycoprotein (S) gene, S-13-4 allele,

partial cds.

ACCESSION AF328994

VERSION AF328994.1 GI:13345390

KEYWORDS

SOURCE

ORGANISM

Arabidopsis lyrata.

Arabidopsis lyrata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosoids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 942)

Schierup,M.H., Mable,B.K., Awadalla,P. and Charlesworth,D.

Identification and Characterization of a Polymorphic Receptor

Kinase Gene Linked to the Self-Incompatibility Locus of Arabidopsis

lyrata

Genetics 158 (1), 387-399 (2001)

21231627

11333247

REFERENCE 2 (bases 1 to 942)

Schierup,M.H., Mable,B.K., Awadalla,P. and Charlesworth,D.

Direct Submission

Submitted (15-DEC-2000) Genetics and Ecology, University of Aarhus,

Ny Munkegade, Building 540, Aarhus 8000, Denmark

Location/Qualifiers

1..942

/organism="Arabidopsis lyrata"

/db\_xref="taxon:59689"

<1..>942

/gene="S"

/product="putative S-glycoprotein"

<1..>942

/gene="S"

<1..>942

/allele="S-13-4"

<1..>942

/gene="S"

/codon\_start=3

/product="putative S-glycoprotein"

/protein\_id="AAK1915.1"

/db\_xref="GI:13345391"

/translation="GNLVIGHSDKLVWSTNLTSGNVRSLVVAELLANGNFMRYSSK

DOGGLWQSFDPDTTLPQMKLGDRKTKLNRLRSWKSSYDPSGGNFTLETRGF

PEEFRLKTDIPHRSGPDGIRISGPEERQVDMVYNFTEDREVAYTFLVTHSIY

SLRTHSYSGYFRFTWTFPANGRWQLWSAPMDLQCDLYPCGAYAYCDTNPPLNCNI

RGFNPSMEQNMWMDGTSGCVRRTPLSCRRDGFLEPMKKMLPTTTMTATVDRRISGKEC

KOKCLMDCNCTAYANADIKNGLGCVIWTGELVDIRT"

260 a 182 c 249 g 251 t

BASE COUNT 260 a 182 c 249 g 251 t

ORIGIN

## Query Match

Best Local Similarity 82.7%; Score 12.4; DB 8; Length 942;

Matches 10; Conservative 71.4%; Pred. No. 2e+03;

Mismatches 4; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 302 TCGAGTTTCGAGTT 289

Search completed: February 25, 2002, 18:00:08  
Job time: 18416 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:21 ; Search time 716.55 Seconds  
(without alignments)  
17.947 Million cell updates/sec

Title: US-09-698-903B-4  
Perfect score: 15  
Sequence: 1 ntgcastwtsgwgtt 15  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930521 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12.4	82.7	15	AAQ79175	Non-specific prime
2	12.4	82.7	15	AAA88455	PCR primer CA54 us
3	12.4	82.7	15	AAA48993	PCR primer MDB285
4	12.4	82.7	15	AAZ59438	PCR primer ARB1 fo
5	12.4	82.7	15	AAF90555	Degenerate primer
6	12.4	82.7	15	AAQ06993	PCR primer MDB285
7	12.4	82.7	15	AAF66681	Degenerate TAIL-PC
8	12.4	82.7	15	AAA89358	Arabidopsis DNA pr
9	12.4	82.7	15	AAC86474	Maize Mac2 gene de
c 10	12.4	82.7	30	AAC84167	Human pollinosis-a
11	12.4	82.7	304	AAH21085	Parallel detection

12	12.4	82.7	335	21	AAQ02398	Human secreted pro
13	12.4	82.7	336	20	AAQ51486	Human secreted pro
c 14	12.4	82.7	388	20	AAQ20974	Polynucleotide seq
15	12.4	82.7	415	22	AAQ06997	Right (5') border
16	12.4	82.7	633	22	AAH21072	Parallel detection
17	12.4	82.7	679	22	AAH59881	Human protein enco
18	12.4	82.7	802	21	AAQ64159	Human pollinosis-a
19	12.4	82.7	836	20	AAH81127	DNA sequence upstr
20	12.4	82.7	951	19	AAQ03315	Pyrolabus fumarius
21	12.4	82.7	1321	22	AAC86489	Maize Mac2n-3 codi
22	12.4	82.7	2914	22	AAH17685	Human cDNA sequenc
23	12.4	82.7	3198	21	AAH48642	Arabidopsis thalia
24	12.4	82.7	4233	20	AAH87946	Candida tropicalis
25	12.4	82.7	4271	21	AAH64160	Human pollinosis-a
26	12.4	82.7	4345	22	AAH99730	Human protein enco
c 27	12.4	82.7	5098	20	AAQ24984	E. coli MG1655 rin
c 28	12.4	82.7	49272	19	AAV35000	Mycobacteriophage
29	11.4	76.0	45	20	AAZ24665	Oligonucleotide fo
c 30	11.4	76.0	165	22	AAI55830	Probe #24516 used
31	11.4	76.0	249	19	AAV03550	DNA sequence that
32	11.4	76.0	438	21	AAC52235	Arabidopsis thalia
33	11.4	76.0	462	22	AAH52999	S. epidermidis ope
34	11.4	76.0	482	21	AAC38143	Zea mays DNA fragm
35	11.4	76.0	489	21	AAV79427	Eucalyptus grandis
c 36	11.4	76.0	503	21	AAC37052	Arabidopsis thalia
c 37	11.4	76.0	539	21	AAQ45066	Human secreted exp
c 38	11.4	76.0	574	21	AAQ09195	Fusarium venenatum
c 39	11.4	76.0	576	16	AAQ83845	Hepatitis C virus
40	11.4	76.0	576	17	AAH16559	Hepatitis C virus
41	11.4	76.0	579	21	AAH10353	Fusarium venenatum
42	11.4	76.0	581	22	AAH29686	Drosophila melanog
c 43	11.4	76.0	590	22	AAI42770	Probe #11456 used
44	11.4	76.0	606	22	AAH07816	Human cDNA clone (
45	11.4	76.0	619	21	AAH48567	Arabidopsis thalia

ALIGNMENTS

RESULT 1  
AAQ79175  
ID AAQ79175 standard; DNA; 15 BP.  
XX AC AAQ79175;  
XX DT 22-JUN-1995 (first entry)  
XX DE Non-specific primer #1, for amplification of T-DNA.  
XX KW Polymerase chain reaction; PCR; amplify; primer; specific;  
KW non-specific; DNA polymerase; deoxyribonucleotide; cyclisation;  
XX ligation; ss.  
OS Synthetic.  
XX PN JP06253843-A.  
XX PD 13-SEP-1994.  
XX PF 03-MAR-1993; 93JP-0043006.  
XX PR 03-MAR-1993; 93JP-0043006.  
XX PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.  
XX PA (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.  
XX DR WPI; 1994-328991/41.  
XX PT New DNA amplification method - without cyclisation of the DNA or  
XX ligation of oligonucleotide(s)  
XX PS Example 1; Page 10; 14pp; Japanese.  
XX

CC The sequences given in AAQ79172-74 are specific primers which were  
 CC used in conjunction with the non-specific primers given in AAQ79175-77  
 CC in the DNA amplification method of the invention. The method  
 CC comprises mixing a DNA sequence and a primer, DNA polymerase and  
 CC deoxyribonucleotides. The DNA is denatured allowing annealing of  
 CC the primers. The amplification cycle consists of one cycle at a  
 CC temperature to allow non-specific annealing of non-specific primers,  
 CC numerous cycles at a temp. to allow specific annealing of specific  
 CC primers and numerous cycles at a temp to allow non-specific primers  
 CC to hybridise specifically to complementary sequences. This method  
 CC allows amplification of neighbouring sequences with known sequences,  
 CC efficiently without cyclisation of the DNA or ligation of  
 CC oligonucleotides.  
 XX  
 SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 15; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 |||||  
 DB 2 tcgastwtsgwgtt 15

RESULT 2  
 AAA88455  
 ID AAA88455 standard; DNA; 15 BP.  
 XX  
 AC AAA88455;  
 XX  
 DT 09-JAN-2001 (first entry)  
 XX  
 DE PCR primer CA54 used to identify Arabidopsis 16713 gene.  
 XX  
 KW 16713 gene; essential gene; herbicide; screening; transgenic plant;  
 KW Arabidopsis thaliana; PCR primer; ss.  
 XX  
 OS Agrobacterium sp.  
 XX  
 PN WO200053782-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 03-MAR-2000; 2000WO-EP01884.  
 XX  
 PR 05-MAR-1999; 99US-0263385.  
 PR 10-MAY-1999; 99US-0309036.  
 PR 08-JUN-1999; 99US-0327662.  
 PR 29-SEP-1999; 99US-0408196.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Patton DA, Ashby CS, Thomas CR, Mcelver JA, Levin Jz;  
 PI Budziszewski GJ;  
 XX  
 XX WPI; 2000-594330/56.  
 XX  
 XX Novel genes isolated from Arabidopsis coding for proteins essential for  
 PT normal plant development, are useful to discover new herbicides and to  
 PT identify inhibitors that are potential herbicides -  
 XX  
 PS Example 4c; Page 53; 92pp; English.  
 XX  
 XX This degenerate oligonucleotide, termed CA54, is a primer used in  
 CC the border rescue of the 16713 gene (see AAA88448) from Arabidopsis  
 CC thaliana embryo-lethal line #16713. DNA flanking the borders of  
 CC line #16713 was isolated using TAIL PCR. CA54 is 1 of 6 primers  
 CC (see AAA88451-56) used in combination with 2 sets of nested, T-DNA  
 CC specific primers for the right border (see AAA88457-59) as well as  
 CC for the left border (see AAA88460-62) of the T-DNA region of pCSA104.

CC The 16713 gene encodes a protein (see AAB19528) that is essential for  
 CC normal plant development. Chemicals that inhibit the function of  
 CC the 16713-encoded protein are likely to have detrimental effects on  
 CC plants and are potentially good herbicide candidates.  
 XX  
 SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 |||||  
 DB 2 tcgastwtsgwgtt 15

RESULT 3  
 AAA48993  
 ID AAA48993 standard; DNA; 15 BP.  
 XX  
 AC AAA48993;  
 XX  
 DT 28-NOV-2000 (first entry)  
 XX  
 DE PCR primer MDB285 used to locate bar gene insertion in rice plant.  
 XX  
 KW Herbicide; phosphinothricin; glufosinate; bialaphos; bar;  
 KW phosphinothricin acetyl transferase; PAT; PCR primer; TAIL;  
 KW thermal asymmetric interlaced; rice; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200026345-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 03-NOV-1999; 99WO-US25667.  
 XX  
 PR 03-NOV-1998; 98US-0185244.  
 XX  
 PA (PLBZ ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Michiels F, Johnson K;  
 XX  
 XX WPI; 2000-365598/31.

XX New glufosinate tolerant rice plants with genomic DNA capable of  
 PT yielding at least 3 restriction fragments and having bar gene under  
 PT control of CaMV 35S promoter at specific location -  
 XX  
 PS Example 3; Page 25; 44pp; English.  
 XX  
 XX The present invention relates to a transgenic glufosinate tolerant rice  
 CC plant. The tolerant rice plants are generated by transformation with  
 CC the Streptomyces hygroscopicus bar gene. This gene encodes the enzyme  
 CC phosphinothricin acetyl transferase (PAT), which confers resistance to  
 CC the herbicidal compounds phosphinothricin (also called glufosinate) and  
 CC bialaphos. The 3' insertion site of the transgene was determined by  
 CC thermal asymmetric interlaced (TAIL) PCR. The present sequence is the  
 CC arbitrary degenerate PCR primer, MDB285, used for the TAIL PCR.  
 CC Glufosinate tolerant rice plants are useful as they give better yields  
 CC and growth, compared to untransformed plants, after application of a  
 CC glufosinate herbicide.  
 XX  
 SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

Db 2 tcgastwtsgwgtt 15  
|||||

RESULT 4

AAZ99438  
ID AAZ99438 standard; DNA; 15 BP.  
XX  
AC AAZ99438;  
XX  
DT 03-JUL-2000 (first entry)  
XX  
DE PCR primer ARB1 for the soybean AX5 promoter.  
XX  
KW Gibberellic acid; copalyl diphosphate synthase; 3beta-hydroxylase;  
KW 2-oxidase; phytoene synthase; C-20 oxidase; 2beta,3beta-hydroxylase;  
KW seed germination; seedling growth; gibberellin biosynthetic pathway;  
KW transgenic plant; hypocotyl; epicotyl; AX5 promoter; PCR primer; ss.  
XX  
OS Glycine max.  
XX  
PN WO200009722-A2.  
XX  
PD 24-FEB-2000.  
XX  
PF 10-AUG-1999; 99WO-US18066.  
XX  
PR 10-AUG-1998; 98US-0096111.  
PR 07-JUN-1999; 99US-0137977.  
XX  
PA (MONS ) MONSANTO CO.  
XX

Brown SM, Elich TD, Heck GR, Kishore GM, Logusch EW, Logusch SJ;  
Piller KJ, Rao S, Ream JE;  
WPI; 2000-224351/19.

Obtaining transgenic plant useful for controlling seed germination and  
seedling growth comprises transgene comprising a sequence expressing  
altered levels of an essential hormone -  
Example 7; Page 98; 267pp; English.

PCR primers AAZ99438-42 were used to amplify the soybean AX5 promoter.  
The amplified sequence is used in the method of the invention. The  
specification describes methods for the inhibition and control of  
gibberellic acid levels. Gibberellic acid levels may be inhibited or  
controlled by use of a chimeric expression construct expressing a RNA  
or protein which suppresses the gibberellin biosynthetic pathway  
sequence, diverts substrate from the pathway, or degrades pathway  
substrates or products. The methods uses copalyl diphosphate synthase,  
3beta-hydroxylase, 2-oxidase, phytoene synthase, C-20 oxidase, and  
a 2beta,3beta-hydroxylase polynucleotides to achieve this. The method  
is used to control seed germination and seedling growth especially to  
regulate gene products of gibberellin biosynthetic pathway and  
restoration of normal seed germination, in transgenic plants. The  
plants produced are gibberellin deficient, and have shortened hypocotyl  
and/or epicotyl phenotypes compared to normal plants.

Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||

Db 2 tcgastwtsgwgtt 15

RESULT 5

AAF90555  
ID AAF90555 standard; DNA; 15 BP.  
XX  
AC AAF90555;  
XX  
DT 22-AUG-2001 (first entry)  
XX  
DE Degenerate primer CA54 used to prime Arabidopsis genomic DNA.  
ET1158 gene; GT6839 gene; ET5262 gene; herbicide; screening;  
herbicide tolerance; transgenic plant; crop protection; PCR primer;  
ss.  
XX  
OS Synthetic.  
XX  
PN WO200144277-A2.  
XX  
PD 21-JUN-2001.  
XX  
PF 14-DEC-2000; 2000WO-EP12748.  
XX  
PR 16-DEC-1999; 99US-0465040.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Wegrich Glover L, Budziszewski GJ, Levin JZ, Zhou Q;  
DR WPI; 2001-398122/42.  
XX  
PT New herbicide target genes encoding proteins having ET1158, GT6839 or  
ET5262 activity, for identifying an inhibitor of protein activity -  
XX  
PS Example 1; Page 39; 67pp; English.  
XX  
CC The present sequence is that of degenerate primer CA54. Arbitrary  
CC degenerate primers LWAD1, CA51, CA52, CA53, CA54 and CA55 (see  
CC AAF90551-56) were used to prime Arabidopsis thaliana genomic DNA  
CC flanking a Ds transposon insertion. The degenerate primers were  
CC used in combination with 2 sets of 3, nested, transposon-specific  
CC primers (see AAF90557-62) that were homologous to regions of the  
CC Ds elements which lie at the outermost ends of the transposons. A  
CC series of low- and high-stringency PCR amplifications were  
CC performed using the TAIL-PCR protocol. DNA fragments were produced  
CC which corresponded to the genomic DNA that was directly adjacent to  
CC the transposon insertion. Sequence analysis of PCR products from  
CC tagged seedling lethal lines ET1158, GT6839 and ET5262 identified 3  
CC novel genes (see AAF90548-50) each of which was essential for  
CC Arabidopsis seedling growth and development. The essentiality of  
CC the genes provides a means of discovering new herbicides. Screening  
CC assays for identifying inhibitors that are potential herbicides are  
CC provided. The invention is also applied to the development of  
CC herbicide tolerant plants, and plant tissues, seeds and cells.  
XX  
SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||

Db 2 tcgastwtsgwgtt 15

RESULT 6

AAZ06993  
ID AAZ06993 standard; DNA; 15 BP.  
XX  
AC AAZ06993;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE PCR primer MDB285 to generate the flanking region of elite event MS-B2.



PD 04-JAN-2001.  
XX  
XX  
PF 21-JUN-2000; 2000WO-EP05761.  
XX  
PR 23-JUN-1999; 99GB-0014623.  
XX  
XX (NOVS ) NOVARTIS AG.  
FA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
XX  
XX  
PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;  
XX  
XX WPI; 2001-137952/14.  
XX  
XX Novel gene encoding a protein that controls gene silencing, in  
PT particular silencing of plant genes -  
XX  
XX Example 4; Page 11; 48pp; English.  
XX  
XX This is the sequence of primer AD1, which is 1 of 7 degenerate  
CC primers (see AA89358-64) expected to bind to Arabidopsis thaliana  
CC DNA flanking the site of a T-DNA insert. These primers were used  
CC with 3 specific, nested primers (see AA89355-57), located close to  
CC the right border of T-DNA, in the cloning of the silencing gene  
CC (see AA89353) of Arabidopsis thaliana. Genomic DNA from the plant  
CC containing only T-DNA co-segregating with a hygromycin resistant  
CC mutant phenotype was isolated. Gene silencing is useful as a  
CC molecular tool for regulating gene expression.  
XX  
XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
SQ

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 2 tcgastwtsgwgtt 15  
Db 2 tcgastwtsgwgtt 15  
|||||  
RESULT 9  
AAC86474  
ID AAC86474 standard; DNA; 15 BP.  
XX  
XX  
AC AAC86474;  
XX  
XX  
DT 01-MAR-2001 (first entry)  
XX  
XX Maize Mac2 gene degenerate PCR primer AD1.  
DE  
XX Tapetum-specific promoter; artificial male sterility; agriculture;  
KW crop yield; disease resistance; AMS; maize; MAC2; PCR primer; ss.  
XX  
XX  
OS Zea mays.  
XX  
XX WO200068403-A2.  
FN  
XX  
XX 16-NOV-2000.  
PD  
XX  
XX 10-MAY-2000; 2000WO-GB01789.  
PF  
XX  
XX 10-MAY-1999; 99GB-0010796.  
PR  
XX  
XX (BT0G-) BIOGENMA UK LTD.  
PA  
XX Paul W, Scott RJ, Hird D, Hodge R;  
PI  
XX WPI; 2001-016099/02.  
DR  
XX  
XX Novel nucleic acid sequences encoding tapetum specific promoters, pMAC2  
PT and pMAC20, or promoter sequences that control the expression of coding  
PT sequence homologous to the promoters, for artificial male sterility -  
XX

PS Example 2; Page 16; 45pp; English.  
XX  
XX The present invention provides tapetum-specific promoters which can be  
CC used to create artificial male sterility (AMS) systems in plants. These  
CC are useful in agriculture as they enable the production of plants with  
CC increased yield and greater disease resistance. They also result in  
CC uniform offspring, which have the same germination time, height of  
CC growth, susceptibility to disease, flowering time etc.  
XX  
XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
SQ

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 2 tcgastwtsgwgtt 15  
Db 2 tcgastwtsgwgtt 15  
|||||  
RESULT 10  
AAC64167/c  
ID AAC64167 standard; DNA; 30 BP.  
XX  
XX AAC64167;  
AC  
XX  
XX 21-FEB-2001 (first entry)  
DT  
XX  
XX Human pollinosis-associated gene 581 hybridisation probe, SEQ ID NO:8.  
DE  
XX  
XX Human; pollinosis-associated gene 581; IgE; immunoglobulin E;  
KW cedar pollen allergy; T-cell; reduced expression; detection;  
KW diagnosis; drug screening; allergic disease; expression analysis;  
KW hybridisation probe; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200065048-A1.  
PN  
XX  
XX 02-NOV-2000.  
PD  
XX  
XX 26-APR-2000; 2000WO-JP02732.  
PF  
XX  
XX 27-APR-1999; 99JP-0120492.  
PR  
XX  
XX (GENO-) GENOX RES INC.  
PA  
XX  
XX Nagasu T, Sugita Y, Kashiwabara T, Oshida T, Obayashi M, Gunji S;  
PI Obayashi I, Imai Y, Yoshida N, Ogawa K, Matsui K;  
XX  
XX WPI; 2000-687341/67.  
DR  
XX  
XX Pollenosis-associated gene 581 undergoing significantly low expression  
PT in subjects with high cedar pollen-specific IgE levels, useful in  
PT diagnosis of allergic diseases and screening drug candidates -  
XX  
XX  
XX Example 8; Page 24; 69pp; Japanese.  
FS  
XX  
XX The invention relates to the human pollinosis-associated gene 581 which  
CC exhibits significantly reduced expression in the T-cells of individuals  
CC with high cedar pollen-specific IgE (immunoglobulin E) levels. The gene  
CC was isolated from T-cells from individuals allergic to cedar pollen using  
CC the differential display method. The invention also relates also relates  
CC to the protein encoded by pollinosis-associated gene 581; to expression  
CC constructs and host cells comprising pollinosis-associated gene 581  
CC nucleic acids; pollinosis-associated gene 581 primers and probes;  
CC antibodies against the protein encoded by the gene; methods of detection  
CC of pollinosis-associated gene 581 nucleic acids; and a method of  
CC diagnosis of allergic diseases via the detection of pollinosis-associated  
CC gene 581 nucleic acids. The invention additionally encompasses methods of  
CC screening drug candidates for the treatment of allergic disease by  
CC measuring the expression of pollinosis-associated gene 581 in pollen

CC antigen-stimulated T-cells in the presence of a test compound relative  
 CC to a control. Pollinosis-associated gene 581 is useful in the diagnosis  
 CC of allergic diseases and in the screening of drug candidates for the  
 CC treatment of such diseases. The present sequence represents a  
 CC hybridisation probe used in human pollinosis-associated gene 581  
 CC expression analysis.

XX Sequence 30 BP; 9 A; 9 C; 4 G; 8 T; 0 other;

Query Match 82.7%; Score 12.4; DB 21; Length 30;  
 Best Local Similarity 71.4%; Pred. No. 3.2e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 ||||:|:|:|  
 Db 30 TCGAGTATGGTGT 17

## RESULT 11

AAH21085  
 ID AAH21085 standard; DNA; 304 BP.  
 XX  
 AC AAH21085;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE Parallel detection of methylated genomic DNA associated DNA SEQ ID 31.  
 XX  
 DE DNA methylation; parallel detection; 5-unmethylated cytosine; CpG;  
 KW CpnPG; amplification; transcription regulation; genetic imprinting;  
 KW tumorigenesis; primer; ss.  
 KW  
 OS Unidentified.

WO200142493-A2.  
 XX  
 PN 14-JUN-2001.

XX  
 PF 06-DEC-2000; 2000WO-DE04381.  
 XX  
 PR 06-DEC-1999; 99DE-1059691.

XX (EPIG-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C;

XX WPI; 2001-381705/40.  
 DR  
 XX

XX Parallel detection of the methylation pattern of many genomic DNA  
 PT regions, useful for detecting aberrant methylation, includes multiple  
 PT amplification of chemically modified DNA -

PS Disclosure; Page 63; 63pp; German.

XX This invention describes a novel method for the parallel detection of the  
 CC methylation status of genomic DNA (1) which involves a (1) sample being  
 CC treated chemically to convert 5-unmethylated cytosine to uracil,  
 CC thymidine or some other base having hybridization behavior different from  
 CC that of C, then amplifying simultaneously at least 10 different fragments  
 CC (of fewer than 2 kb) using synthetic oligonucleotide (ON) primers. These  
 CC primers are based on regulatory, transcribed and/or translated segments  
 CC present in the sample after chemical treatment. The sequence context of  
 CC all, or some, of the CpG and CpNpG motifs in the amplified products is  
 CC then determined. The method is used to detect aberrant methylation  
 CC patterns in the genome, these are implicated in regulation of  
 CC transcription, genetic imprinting and tumorigenesis. Many target regions  
 CC in the genome can be analyzed simultaneously and it is not essential to  
 CC know the sequence context of all targeted regions. Primers may be  
 CC designed for preferential amplification of particular segments of  
 CC interest (e.g. promoters and exons).

XX Sequence 304 BP; 37 A; 35 C; 131 G; 101 T; 0 other;

Query Match 82.7%; Score 12.4; DB 22; Length 304;  
 Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 ||||:|:|:|  
 Db 112 tcgagtttggagtt 125

## RESULT 12

AAC02398  
 ID AAC02398 standard; cDNA; 335 BP.  
 XX  
 AC AAC02398;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 2396.  
 XX  
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 KW  
 OS Homo sapiens.

XX EP1033401-A2.  
 XX  
 PD 06-SEP-2000.

XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.  
 DR P-PSDB; AAG02392.  
 XX  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 2396; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.

XX Sequence 335 BP; 71 A; 75 C; 89 G; 89 T; 11 other;  
 SQ

Query Match 82.7%; Score 12.4; DB 21; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 ||||:|:|:|  
 Db 194 tcgagtttggagtt 207

```

RESULT 13
AAAX51486
ID AAX51486 standard; cDNA: 336 BP.
XX
AC AAX51486;
XX
DT 21-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO:65.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906549-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01231.
XX
PR 01-AUG-1997; 97US-0905279.
XX
PA (GEST ) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153779/13.
XX
DR P-PSDB; AAY12708.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS Claim 1; Page 192; 522pp; English.
XX
CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12681 to
CC AAY12913, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 336 BP; 72 A; 75 C; 89 G; 89 T; 11 other;

Query Match 82.7%; Score 12.4; DB 20; Length 336;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
Db 195 tcgagtttgaggtt 208
||||:|:|:|:|

RESULT 14
AAAX20974/c
ID AAX20974 standard; DNA; 388 BP.

```

```

XX AAX20974;
XX AC
XX DT 05-MAY-1999 (first entry)
XX DE
XX DE Polynucleotide sequence from the genome of Treponema pallidum.
XX KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX KW enzyme production; ds.
XX OS Treponema pallidum.
XX PN WO9859034-A2.
XX PD 30-DEC-1998.
XX PF 23-JUN-1998; 98WO-US13041.
XX PR 24-JUN-1997; 97US-0050667.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Fraser CM;
XX DR WPI; 1999-081273/07.
XX PT New isolated Treponema pallidum nucleic acids - used to develop
XX PT products for the detection, diagnosis, characterisation, prevention
XX PT and therapy of T. pallidum infections, particularly syphilis
XX PS Claim 1; Page 998-999; 1150pp; English.
XX CC AAX20500-21243 represent polynucleotide sequences from the genome of
XX CC Treponema pallidum. The sequences can be used for detection
XX CC diagnosis, characterisation, prevention and therapy for T. pallidum
XX CC infections, particularly syphilis. They can also be used for detecting
XX CC diseases related to Borrelia infections in animals, and for the
XX CC production of biosynthetic products such as enzymes.
XX SQ Sequence 388 BP; 121 A; 68 C; 77 G; 117 T; 5 other;

Query Match 82.7%; Score 12.4; DB 20; Length 388;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
Db 323 TCGAGTTTGGTGTT 310
||||:|:|:|:|

RESULT 15
AAD06997
ID AAD06997 standard; DNA; 415 BP.
XX
AC AAD06997;
XX
DT 06-AUG-2001 (first entry)
XX
XX Right (5') border flanking region of elite event MS-B2.
XX
XX MS-B2 elite event; transgenic Brassica plant; transformation event;
XX KW male-sterility gene; ds.
XX OS Chimeric - Agrobacterium sp.
XX OS Chimeric - Brassica sp.
XX
XX Key Location/Qualifiers
XX FT misc_feature 1..234
XX FT /*tag= a
XX FT /note= "Corresponds to plant DNA"
XX FT 235..415
XX FT misc_feature
XX FT /*tag= b

```

/note= "Corresponds to T-DNA"

FT WO200131042-A2.  
 XX  
 XX  
 XX PD 03-MAY-2001.  
 XX  
 XX PF 26-OCT-2000; 2000WO-EF10680.  
 XX  
 XX PR 29-OCT-1999; 99US-0430497.  
 XX  
 XX PA (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 XX PI Weston B, De Beuckeleer M;  
 XX  
 XX DR WPI; 2001-300517/31.  
 XX  
 XX PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX  
 XX PS Claim 11; Page 51; 53pp; English.  
 XX  
 CC The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is right (5') border flanking region of elite event  
 CC MS-B2.  
 XX  
 SQ Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 82.7%; Score 12.4; DB 22; Length 415;  
 Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
 |||||:|:|:|  
 Db 2 tcgagtttggtt 15

Search completed: February 25, 2002, 18:17:22  
 Job time: 16680 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:11 ; Search time 301.6 Seconds  
(without alignments)  
11.264 Million cell updates/sec

Title: US-09-698-903B-4  
Perfect score: 15  
Sequence: 1 ntcgastwtsgwgtt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCRU5\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12.4	82.7	15	4	US-09-503-391-13
c 2	12.4	82.7	49272	1	US-08-614-770A-1
3	11.4	76.0	576	1	US-08-086-428B-1
4	11.4	76.0	576	2	US-08-468-570-1
5	11.4	76.0	576	2	US-08-290-665A-1
6	11.4	76.0	576	5	PCT-US95-10398-1
7	11.4	76.0	1268	4	US-09-046-894-50
c 8	11.4	76.0	2220	5	PCT-US95-13749-2
c 9	11.4	76.0	2341	3	US-09-187-049-11
10	11.4	76.0	2493	3	US-08-804-439A-11
11	11.4	76.0	2493	3	US-08-720-229-11
12	11.4	76.0	2652	2	US-08-953-492-1
13	11.4	76.0	2873	1	US-08-149-695-1
14	11.4	76.0	2873	1	US-08-377-228-1
15	11.4	76.0	3139	2	US-08-693-457-1
16	11.4	76.0	3139	4	US-09-265-731-1
17	11.4	76.0	12311	4	US-08-750-717-1
c 18	11.2	74.7	1567	1	US-08-181-271A-100
c 19	11.2	74.7	1567	1	US-08-449-315-100
c 20	11.2	74.7	1567	1	US-08-444-803-100
c 21	11.2	74.7	1567	1	US-08-449-043-100
c 22	11.2	74.7	1567	1	US-08-456-265A-100
c 23	11.2	74.7	1567	1	US-08-455-416-100
c 24	11.2	74.7	1567	1	US-08-455-244-100
c 25	11.2	74.7	1567	1	US-08-454-876-100
c 26	11.2	74.7	1567	2	US-08-457-364-100
c 27	11.2	74.7	1567	2	US-08-456-362-100

c 28	11.2	74.7	1567	2	US-08-456-240-100	Sequence 100, App
c 29	11.2	74.7	1567	2	US-08-455-736-100	Sequence 100, App
c 30	11.2	74.7	1567	2	US-08-971-217-100	Sequence 100, App
c 31	11.2	74.7	1567	4	US-09-350-600-100	Sequence 100, App
32	11.2	74.7	2068	1	US-08-309-341-1	Sequence 1, Appli
33	11.2	74.7	2068	1	US-08-608-267-1	Sequence 1, Appli
34	11.2	74.7	2068	1	US-08-608-452-1	Sequence 1, Appli
35	11.2	74.7	2068	1	US-08-608-224-1	Sequence 1, Appli
36	11.2	74.7	2068	2	US-08-967-149-1	Sequence 1, Appli
c 37	11.2	74.7	3117	1	US-08-172-331B-3	Sequence 3, Appli
c 38	11.2	74.7	3280	1	US-08-259-000-4	Sequence 4, Appli
c 39	11.2	74.7	3280	1	US-08-729-767-6	Sequence 6, Appli
c 40	11.2	74.7	4643	2	US-08-605-106-6	Sequence 6, Appli
41	11.2	74.7	7400	1	US-07-674-852-1	Sequence 1, Appli
42	11.2	74.7	7400	3	US-08-473-185-1	Sequence 1, Appli
43	11.2	74.7	7400	4	US-09-171-387-3	Sequence 3, Appli
c 44	11.2	74.7	12494	4	US-08-935-312-13	Sequence 13, Appli
c 45	11.2	74.7	12494	4	US-08-848-760B-33	Sequence 33, Appli

ALIGNMENTS

RESULT 1  
US-09-503-391-13  
; Sequence 13, Application US/09503391  
; Patent No. 6300091  
; GENERAL INFORMATION:  
; APPLICANT: Patton, David A.  
; APPLICANT: Ashby, Carl S.  
; APPLICANT: Thomas, Carla R.  
; APPLICANT: McElver, John A.  
; APPLICANT: Budziszewski, Gregory J.  
; APPLICANT: Levin, Joshua Z.  
; TITLE OF INVENTION: Herbicide Target Genes and Methods  
; FILE REFERENCE: PB/5-30852A  
; CURRENT APPLICATION NUMBER: US/09/503,391  
; CURRENT FILING DATE: 2000-02-14  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: oligonucleotide  
US-09-503-391-13

Query Match 82.7%; Score 12.4; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 69; Indels 0; Gaps 0;  
Matches 14; Conservative 0; Mismatches 0

Qy 2 tcgastwtsgwgtt 15  
| | | | | | | | | | | | | | |  
Db 2 tcgastwtsgwgtt 15

RESULT 2  
US-08-614-770A-1/c  
; Sequence 1, Application US/08614770A  
; Patent No. 5773267  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL  
; TITLE OF INVENTION: D29 SHUTTLE GRAHMS AND USES THEREOF  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
; STREET: 90 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.

ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE  
MEDIUM TYPE: DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/614,770A  
FILING DATE: MARCH 7, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ELIZABETH A. BOGOSIAN  
REGISTRATION NUMBER: 39,911  
REFERENCE/DOCKET NUMBER: 96700/402  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 49272  
TYPE: NUCLEIC ACID  
STRANDEDNESS: DOUBLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: OLIGONUCLEOTIDE  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: MYCOBACTERIOPHAGE  
INDIVIDUAL ISOLATE: D29  
US-08-614-770A-1

Query Match 82.7%; Score 12.4; DB 1; Length 49272;  
Best Local Similarity 71.4%; Pred. No. 1e+02; Indels 0; Gaps 0;  
Matches 10; Conservative 4; Mismatches 0;

QY 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 29087 TCGAGTTCGTGT 29074

RESULT 3  
US-08-086-428B-1  
; Sequence 1, Application US/08086428B  
; Patent No. 5514539  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R. H. AND  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 159  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/086,428B  
; FILING DATE: 29-JUN-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4070  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 576 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; ORIGINAL SOURCE:  
; ORGANISM: homosapiens  
; INDIVIDUAL ISOLATE: DK7  
; US-08-086-428B-1

Query Match 76.0%; Score 11.4; DB 1; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgt 14

||||:|:|:|

Db 55 TCGAGTTCGTGT 67

RESULT 4  
US-08-468-570-1  
; Sequence 1, Application US/08468570  
; Patent No. 5871962  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R. H. AND  
; APPLICANT: PURCELL, R. H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 159  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,570  
; FILING DATE: 6-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/086,428  
; FILING DATE: 29-JUN-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4070US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 576 base pairs  
; TYPE: nucleic acid

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; ORIGINAL SOURCE:  
; ORGANISM: homosapiens  
; INDIVIDUAL ISOLATE: DK7  
US-08-468-570-1

Query Match 76.0%; Score 11.4; DB 2; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgt 14  
||||:|:|:|

Db 55 TCGAGTATCGTGT 67

RESULT 5  
US-08-290-665A-1  
; Sequence 1, Application US/08290665A  
; Patent No. 5882852  
; GENERAL INFORMATION:  
; APPLICANT: BURKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,665A  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 751-6840  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 576 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; ORIGINAL SOURCE:  
; ORGANISM: homosapiens  
; INDIVIDUAL ISOLATE: DK7  
US-08-290-665A-1

Query Match 76.0%; Score 11.4; DB 2; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgt 14  
||||:|:|:|

Db 55 TCGAGTATCGTGT 67

RESULT 6  
PCT-US95-10398-1  
; Sequence 1, Application PC/TUS9510398  
; GENERAL INFORMATION:  
; APPLICANT: BURKH, J., MILLER, R.H. AND  
; APPLICANT: PURCELL, R.H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE  
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 263  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/10398  
; FILING DATE: 15-AUG-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/086,428  
; FILING DATE: 29 JUNE 1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/290/665  
; FILING DATE: 15 AUGUST 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4116  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 751-6840  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 576 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; ORIGINAL SOURCE:  
; ORGANISM: homosapiens  
; INDIVIDUAL ISOLATE: DK7  
PCT-US95-10398-1

Query Match 76.0%; Score 11.4; DB 5; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgt 14  
||||:|:|:|

Db 55 TCGAGTATCGTGT 67

RESULT 7  
US-09-046-894-30  
; Sequence 30, Application US/09046894  
; Patent No. 6190857  
; GENERAL INFORMATION:  
; APPLICANT: Ralph, David  
; APPLICANT: An, Gang  
; APPLICANT: O'Hara, Mark S.  
US-09-046-894-30

APPLICANT: Veltri, Robert  
TITLE OF INVENTION: DIAGNOSIS OF DISEASE STATE USING mRNA  
TITLE OF INVENTION: PROFILES IN PERIPHERAL LEUKOCYTES  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,894  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,576  
FILING DATE: 24-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakashima, Richard A.  
REGISTRATION NUMBER: P-42,023  
REFERENCE/DOCKET NUMBER: UROC:014  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1268 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-046-894-30

Query Match 76.0%; Score 11.4; DB 4; Length 1268;  
Best Local Similarity 69.2%; Pred. No. 3.5e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 3 cgaatwtsgwgtt 15  
||||:|:|:|

Db 850 CGAGTTTCGTGTT 862

RESULT 8  
PCT-US95-13749-2/c  
Sequence 2, Application PC/TUS9513749  
GENERAL INFORMATION:  
APPLICANT: Amgen Inc.  
TITLE OF INVENTION: COMPOSITIONS FOR INCREASED  
TITLE OF INVENTION: BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13749  
FILING DATE:  
CLASSIFICATION:  
INFORMATION FOR SEQ ID NO: 2:  
.

SEQUENCE CHARACTERISTICS:  
LENGTH: 2220 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 536..1024  
PCT-US95-13749-2

Query Match 76.0%; Score 11.4; DB 5; Length 2220;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgt 14  
||||:|:|:|

Db 975 TCGAGTTTGAGT 963

RESULT 9  
US-09-187-049-11/c  
Sequence 11, Application US/09187049  
Patent No. 6117666  
GENERAL INFORMATION:  
APPLICANT: Lamppa, Gayle K.  
TITLE OF INVENTION: PLASTID PROTEOLYTIC PROCESSING ENZYME  
TITLE OF INVENTION: THAT CLEAVES PRECURSOR POLYPEPTIDES  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BRINKS HOFER GILSON & LIONE  
STREET: P.O. Box 10395  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/187,049  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/695,177  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Martin, Alice O.  
REGISTRATION NUMBER: 35,601  
REFERENCE/DOCKET NUMBER: 7814/16  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312 321-4200  
TELEFAX: 312 321-4299  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2341 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Arabidopsis CPE  
US-09-187-049-11

Query Match 76.0%; Score 11.4; DB 3; Length 2341;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

Qy 2 tcgastwtsgwt 14
Db 599 TCGACTATGGAGT 587

RESULT 10
US-08-804-439A-11
; Sequence 11, Application US/08804439A
; Patent No. 6015565
; GENERAL INFORMATION:
; APPLICANT: Rose, Timothy M.
; APPLICANT: Bosch, Marnix L.
; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Ste 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,439A
; FILING DATE: February 21, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09176/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-804-439A-11

Query Match 76.0%; Score 11.4; DB 3; Length 2493;
Best Local Similarity 69.2%; Pred. No. 3.6e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14
Db 1369 TCGAGTATCGTGT 1381

RESULT 11
US-08-720-229-11
; Sequence 11, Application US/08720229
; Patent No. 6022542
; GENERAL INFORMATION:
; APPLICANT: Rose, Timothy M.
; APPLICANT: Bosch, Marnix L.
; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road

```

```

; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/720,229
; FILING DATE: 26-SEP-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Schiff, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20002.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-720-229-11

Query Match 76.0%; Score 11.4; DB 3; Length 2493;
Best Local Similarity 69.2%; Pred. No. 3.6e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14
Db 1369 TCGAGTATCGTGT 1381

RESULT 12
US-08-953-492-1
; Sequence 1, Application US/08953492
; Patent No. 5849555
; GENERAL INFORMATION:
; APPLICANT: Brown, James
; APPLICANT: Jaworski, Deborah
; APPLICANT: Lawlor, Elizabeth
; APPLICANT: Wang, Min
; TITLE OF INVENTION: NOVEL vals
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,492
; FILING DATE: 17-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/844,064
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9607991.8
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:

```

NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31458-4/1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2652 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-953-492-1

Query Match 76.0%; Score 11.4; DB 2; Length 2652;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 tcgastwtsgwgt 14  
||||:|:|:|  
Db 1892 TCGAGTTGGTGT 1904

RESULT 13  
US-08-149-695-1  
Sequence 1, Application US/08149695  
Patent No. 5412085  
GENERAL INFORMATION:  
APPLICANT: Allen, Rebecca L.  
APPLICANT: Lonsdale, David M.  
TITLE OF INVENTION: A Pollen-Specific Promoter From Maize  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAndrews, Held and Malloy  
STREET: 500 W. Madison, 34th Floor  
CITY: Chicago  
STATE: Illinois  
COUNTRY: U.S.A.  
ZIP: 60661  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/149,695  
FILING DATE:

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/911,532  
FILING DATE: 09-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pochopien, Donald J.  
REGISTRATION NUMBER: 32167  
REFERENCE/DOCKET NUMBER: 92 P139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312)707-8889  
TELEFAX: (312)707-9155  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Zea mays  
STRAIN: Line W22

TISSUE TYPE: Pollen  
US-08-149-695-1

Query Match 76.0%; Score 11.4; DB 1; Length 2873;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cgastwtsgwgtt 15  
||||:|:|:|  
Db 111 CGAGTTGGAGTT 123

RESULT 14  
US-08-377-228-1  
Sequence 1, Application US/08377228  
Patent No. 5545546  
GENERAL INFORMATION:  
APPLICANT: Allen, Rebecca L.  
APPLICANT: Lonsdale, David M.  
TITLE OF INVENTION: A Pollen-Specific Promoter From Maize  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/377,228  
FILING DATE: 24-JAN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/149,695  
FILING DATE: 09-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/911,532  
FILING DATE: 09-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 33229/290/PIHI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-377-228-1

Query Match 76.0%; Score 11.4; DB 1; Length 2873;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cgastwtsgwgtt 15  
||||:|:|:|  
Db 111 CGAGTTGGAGTT 123

RESULT 15  
US-08-693-457-1  
Sequence 1, Application US/08693457

; Patent No. 5880330  
; GENERAL INFORMATION:  
; APPLICANT: Weigel et al., Detlef  
; TITLE OF INVENTION: SHOOT MERISTEM SPECIFIC PROMOTER  
; TITLE OF INVENTION: SEQUENCES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/693,457  
; FILING DATE: 07-AUG-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Halle, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07251/012001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3139 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 571..1900  
US-08-693-457-1

Query Match 76.0%; Score 11.4; DB 2; Length 3139;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Oy 3 cgastwtsgwgtt 15  
|||:|:|:|:|  
Db 1546 CGACTTGGAGTT 1558

Search completed: February 25, 2002, 18:05:13  
Job time: 18561 sec

**THIS PAGE BLANK** (USPTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:20:54 ; Search time 8261.74 Seconds  
(without alignments)  
19.510 Million cell updates/sec

Title: US-09-698-903B-4  
Perfect score: 15  
Sequence: 1 ntcgastwtsgwgtt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	12.4	82.7	101	13 CNS04Q4L	AL302142 Tetraodon
C 2	12.4	82.7	203	11 BG199530	RG199530 RSP18921
C 3	12.4	82.7	204	10 BB258732	BB258732 BB258732
C 4	12.4	82.7	210	10 BB411504	BB411504 BB411504
C 5	12.4	82.7	216	10 AL367282	AL367282 MTBA13H05
C 6	12.4	82.7	221	10 AV427766	AV427766 AV427766
C 7	12.4	82.7	226	11 BG057582	BG057582 nah92601.
C 8	12.4	82.7	247	10 AV129451	AV129451 AV129451
C 9	12.4	82.7	248	10 BB074324	BB074324 BB074324
C 10	12.4	82.7	252	10 AV331577	AV331577 AV331577
C 11	12.4	82.7	257	10 BE530468	BE530468 M77D22STM
C 12	12.4	82.7	288	13 AZ921129	AZ921129 1006024D0

13	12.4	82.7	291	10	BB191477
C 14	12.4	82.7	297	10	AV069588
C 15	12.4	82.7	298	11	BF460492
C 16	12.4	82.7	301	10	AJ284234
C 17	12.4	82.7	302	11	F14285
C 18	12.4	82.7	307	10	AV558749
C 19	12.4	82.7	307	10	BB255970
C 20	12.4	82.7	310	10	AV421222
C 21	12.4	82.7	316	10	BB250639
C 22	12.4	82.7	335	11	BF098143
C 23	12.4	82.7	342	10	AV209024
C 24	12.4	82.7	343	10	AV392544
C 25	12.4	82.7	344	10	AW693454
C 26	12.4	82.7	345	11	D40568
C 27	12.4	82.7	349	10	AV534046
C 28	12.4	82.7	355	10	AJ284258
C 29	12.4	82.7	355	10	AV441349
C 30	12.4	82.7	356	11	BF590524
C 31	12.4	82.7	359	10	AV410421
C 32	12.4	82.7	360	11	C71547
C 33	12.4	82.7	366	10	BE321781
C 34	12.4	82.7	374	11	BF778123
C 35	12.4	82.7	376	10	AA754245
C 36	12.4	82.7	376	10	AV202387
C 37	12.4	82.7	376	10	AV202461
C 38	12.4	82.7	377	10	AV422965
C 39	12.4	82.7	382	10	AV409979
C 40	12.4	82.7	382	10	AW796228
C 41	12.4	82.7	386	11	BI241776
C 42	12.4	82.7	394	10	BE192723
C 43	12.4	82.7	397	11	BF483146
C 44	12.4	82.7	405	11	BG630990
C 45	12.4	82.7	406	10	AW859801

#### ALIGNMENTS

RESULT 1  
CNS04Q4L/C

LOCUS  
DEFINITION

CNS04Q4L 101 bp DNA GSS 24-MAY-2000  
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 128H18 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION  
AL302142

VERSION  
AL302142.1 GI:8179769

KEYWORDS  
GSS; genome survey sequence.

SOURCE  
Tetraodon nigroviridis.

ORGANISM  
Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE  
AUTHORS

1 (bases 1 to 101)  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizes, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

Unpublished  
2 (bases 1 to 101)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizes, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

Unpublished  
3 (bases 1 to 101)

Genoscope.  
Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis

COMMENT



cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATCCGCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 49 a 40 c 36 g 79 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 204;  
Best Local Similarity 71.4%; Pred. No. 3e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 4; Mismatches 0;

Qy 2 tcgastwtsgwgtt 15  
||||:|:|:|  
Db 99 TCGACTTTGGTGTT 112

RESULT 4  
LOCUS BB411504/c 210 bp mRNA EST 16-JUL-2000  
DEFINITION BB411504 RIKEN full-length enriched, 7 days embryo Mus musculus cDNA clone C430023B01 3', mRNA sequence.

ACCESSION BB411504  
VERSION BB411504.1 GI:9232859  
KEYWORDS EST.  
SOURCE house mouse.

ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 210)  
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES  
Location/Qualifiers  
1. .210

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="C430023B01"  
/clone\_lib="RIKEN full-length enriched, 7 days embryo"  
/dev\_stage="7 days embryo"  
/lab\_host="DH10B"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCTCGAGTTAATTAATCCGCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATTAATCCGCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 63 a 54 c 41 g 52 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 210;  
Best Local Similarity 71.4%; Pred. No. 3e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 4; Mismatches 0;

Qy 2 tcgastwtsgwgtt 15  
||||:|:|:|  
Db 18 TCGAGTATCGTGTT 5

RESULT 5  
LOCUS AL367282/c 216 bp mRNA EST 03-AUG-2000  
DEFINITION MTBAL13H05F1 MCBA Medicago truncatula cDNA clone MTBAL13H05 T3, mRNA sequence.

ACCESSION AL367282.1 GI:9667035  
VERSION AL367282  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE 1 (bases 1 to 216)  
AUTHORS Journet, E. P., Crespeau, H., van-Tuinen, D., Gouzy, J., Jaillon, O., Niebel, A., Carreau, V., Chatagnier, O., Kahn, D., Gianinazzi-Pearson, V. and Gamas, P.  
Medicago truncatula ESTs from nitrogen-starved roots  
Unpublished (2000)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

CONTACT : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 273126 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).

FEATURES  
Location/Qualifiers  
1. .216  
/organism="Medicago truncatula"  
/cultivar="Jemalong"  
/db\_xref="taxon:3880"  
/clone="MTBAL13H05"  
/clone\_lib="MtBA"  
/tissue\_type="root tips"  
/dev\_stage="harvested after 3 days of N-starvation"  
/note="Vector: pBluescript pSK; Site\_1: EcoRI; Site\_2:

XhoI; Plants were grown in an aeroponic chamber for 14 days on nitrogen-rich medium followed by 3 days on N-free medium. RNA was extracted from root tips (1-3 cm). cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zapXR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequençage (Genoscope, Evry, France)."

69 a 56 c 34 g 56 t 1 others

Query Match 82.7%; Score 12.4; DB 10; Length 216;

Best Local Similarity 71.4%; Pred. No. 3e+03;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

Db 74 TCGAGTTTCGAGTT 61

RESULT 6

AV427766/c

LOCUS

DEFINITION AV427766 Lotus japonicus young plants (two-week old) Lotus japonicus cDNA clone MM086e02\_r 5', mRNA sequence.

ACCESSION AV427766

VERSION AV427766.1

KEYWORDS EST.

SOURCE Lotus japonicus.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

1 (bases 1 to 221)

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus

DNA Res. 7 (2), 127-130 (2000)

20277479

Contact: Yasukazu Nakamura

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: ynakam@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source

1..221

/organism="Lotus japonicus"

/db\_xref="taxon:34305"

/clone="MM086e02\_r"

/dev\_stage="young plants (two-week old)"

/notes="Vector: pBluescriptII SK; Site\_1: EcoRI; Site\_2: XhoI; isolate=MiyakoJima MG-20"

73 a 60 c 25 g 63 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 82.7%; Score 12.4; DB 10; Length 221;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

Db 49 TCGAGTTTCGAGTT 36

RESULT 7

BG057582

LOCUS BG057582 226 bp mRNA EST 25-JAN-2001

DEFINITION nah92d01.x1 NCI\_CGAP\_HN17 Homo sapiens cDNA clone IMAGE:4258200 3', mRNA sequence.

ACCESSION BG057582

VERSION BG057582.1

KEYWORDS GI:12523216

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 226)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

unknown library type

Seq primer: -40UP from Gibco.

Location/Qualifiers

1..226

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4258200"

/clone\_lib="NCI\_CGAP\_HN17"

/tissue\_type="normal epithelium"

/lab\_host="DH10B"

/note="Organ: nasopharynx; Vector: pAMP10; mRNA made from normal nasopharyngeal epithelium, cDNA made by oligo-dT priming. Non-directionally cloned into UDG sites.

Size-selected on agarose gel, average insert size 500 bp.

Primary library. cDNA Library Preparation: David B. Krizman, Ph.D. REFERENCE: Krizman et al. (1996) Cancer Research 56:5380-5383."

63 a 59 c 43 g 61 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 82.7%; Score 12.4; DB 11; Length 226;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

Db 2 TCGACTTTGGTGT 15

RESULT 8

AV129451/c

LOCUS

DEFINITION

AV129451 Mus musculus C57BL/6J 11-day embryo Mus musculus cDNA clone 2700068H03, mRNA sequence.

ACCESSION AV129451

VERSION AV129451.1

KEYWORDS GI:5315686

SOURCE EST.

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

1 (bases 1 to 247)

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Alzawa, K.,

Akai, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara

, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,

Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,

Sato, K., Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Sugahara

, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,

Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,

Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp  
Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

```
FEATURES             Location/Qualifiers
     source           1. .247
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="2700068H03"
                     /clone_lib="Mus musculus C57BL/6J 11-day embryo"
                     /sex="mixed"
                     /dev_stage="11-day embryo"
BASE COUNT          77 a 41 c 34 g 95 t
ORIGIN
```

```
Query Match          82.7%; Score 12.4; DB 10; Length 247;
Best Local Similarity 71.4%; Pred. No. 3.1e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy  2 tcgastwtsgwgtt 15
||||:|:|:|:|
Db  121 TCGAGTTTGGTGTT 108
```

```
RESULT  9
BB074324 248 bp mRNA EST 27-JUN-2000
LOCUS BB074324 RIKEN full-length enriched, adult male diencephalon Mus
DEFINITION musculus cDNA clone 9330013E13 3' similar to S78234 nuc2 homolog,
mRNA sequence.
BB074324
BB074324.1 GI:8584322
EST.
```

```
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 248)
```

```
REFERENCE
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya
T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
```

```
TITLE Contact: Yoshihide Hayashizaki
JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic
COMMENT Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
```

Email: genome-res@rtc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

```
FEATURES             Location/Qualifiers
     source           1. .248
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="9330013E13"
                     /clone_lib="RIKEN full-length enriched, adult male
                     diencephalon"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: Sali; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGATTCGAGTTAATTAATTAATCCGCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
                     BamHI"
```

```
BASE COUNT          68 a 31 c 50 g 99 t
ORIGIN
```

```
Query Match          82.7%; Score 12.4; DB 10; Length 248;
Best Local Similarity 71.4%; Pred. No. 3.1e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy  2 tcgastwtsgwgtt 15
||||:|:|:|:|
Db  197 TCGAGTATGGTGTT 210
```

```
RESULT  10
AV331577
```

```
LOCUS AV331577 252 bp mRNA EST 11-NOV-1999
DEFINITION AV331577 RIKEN full-length enriched, adult male medulla oblongata
Mus musculus cDNA clone 6330524M18 3', mRNA sequence.
```

```
ACCESSION AV331577
VERSION AV331577.1 GI:6371629
```

```
KEYWORDS EST.
SOURCE house mouse.
```

```
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 252)
```

```
REFERENCE
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
```

Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y., Yoshiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al. 1999)

# TITLE JOURNAL COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh,M., Kitsumai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

## FEATURES source

Location/Qualifiers  
1. .252  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="630524M18"  
/clone\_lib="RIKEN full-length enriched, adult male medulla oblongata"  
/sex="male"  
/tissue\_type="medulla oblongata"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTTTTTTTNN 3'], cDNA was GAGAGAGAGATCCAGAGCTCTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATTAATTAATTCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 67 a 53 c 53 g 79 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 252;  
Best Local Similarity 71.4%; Pred. No. 3.2e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
||||:|||||

Db 129 TCGACTTGGTGT 142

## RESULT 11 BE530468/c LOCUS

DEFINITION

BE530468 257 bp mRNA EST 19-MAR-2001  
M77D22STM Arabidopsis developing seed Arabidopsis thaliana cDNA clone 600039982R1 5', mRNA sequence.

ACCESSION

BE530468

VERSION

BE530468.1 GI:9788458

KEYWORDS

EST.

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi

AUTHORS

White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de

TITLE

1 (bases 1 to 257)

JOURNAL

A new set of Arabidopsis expressed sequence tags from developing

MEDLINE

Plant Physiol. 124 (4), 1582-1594 (2000)

COMMENT

20567808

Contact: Benning, C

Dept. of Biochemistry & Molecular Biology

Michigan State University

224 Biochemistry, Michigan State University, East Lansing, MI 48824

USA

Tel: 517 355 1609

Fax: 517 353 9334

Email: benning@msu.edu

Clones were originally prepared at Michigan State University.

Arabidopsis Biological Resource Center, The Ohio State University,

309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210

USA, FAX: 6142920603 TEL: 6142929371.

FEATURES

Location/Qualifiers

1. .257

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="600039982R1"

/clone\_lib="Arabidopsis developing seed"

/tissue\_type="seed"

/dev\_stage="5-13 days after flowering"

/lab\_host="E.coli"

/note="Organ: Developing seed; Vector: pBluescript SK-;

Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 85 a 71 c 35 g 66 t

ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 257;

Best Local Similarity 71.4%; Pred. No. 3.2e+03;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|||||

Db 58 TCGAGTTTCGTGT 45

RESULT 12

AZ921129/c

LOCUS

AZ921129

DEFINITION

1006024D02.y1 1006 - RescueMu Grid G Zea mays genomic, DNA

ACCESSION

AZ921129

VERSION

AZ921129.1 GI:13392455

KEYWORDS

GSS.

SOURCE

Zea mays.

ORGANISM

Zea mays

REFERENCE

1 (bases 1 to 288)

AUTHORS

Walbot,V.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 288)

AUTHORS

Walbot,V.



# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AV069588  
AV069588.1 GI:5189416  
EST.  
house mouse.  
Mus musculus

## REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 297)  
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara  
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,  
Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,  
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara  
Y., Suzuki, H., Suzuki, H., Tatenno, M., Tomaru, Y., Tominaga, N.,  
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,  
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

## TITLE JOURNAL COMMENT

Unpublished (1999)  
Contact: Chie Owa  
Genome Science Laboratory  
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel.: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-resetc.riken.go.jp

Thermostabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length cDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3453-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## FEATURES source

Location/Qualifiers  
1..297  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="201031011"  
/clone\_lib="Mus musculus small intestine C57BL/6J adult"  
/sex="male"  
/tissue\_type="small intestine"  
/dev\_stage="adult"

BASE COUNT 95 a 89 c 68 g 45 t

## ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 297;

Best Local Similarity 71.4%; Pred. No. 3.3e+03;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastvtsgwgtt 15

Db 74 TCGAGTTGGTGT 61

## RESULT 15

BF460492/c

LOCUS

DEFINITION

BF460492 298 bp mRNA EST 04-DEC-2000

UI-M-CG0p-bmd-c-02-0-UI.s1 NIH\_BMAP\_Ret4\_S2 Mus musculus cDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 298)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL

MEDLINE

## COMMENT

Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: m5t5@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
retina tissue cDNA Library Preparation: M.B. Soares Lab Clone  
distribution: Researchers may obtain BMAP cDNA clones from RESEARCH  
GENETICS. It should be noted that Bento Soares is generating a  
small number of additional specialized non-redundant arrays of BMAP  
cDNAs whose availability will be considered under appropriate and  
limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES

source

Location/Qualifiers  
1..298  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CG0p-bmd-c-02-0-UI"  
/clone\_lib="NIH\_BMAP\_Ret4\_S2"  
/lab\_host="PH10B (Life Technologies)"  
/note="Vector: pT7p3D-Pac (Pharmacia) with a modified  
polylinker: Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_Ret4\_S2 library is a subtracted library,  
ultimately derived from mouse retina tissue libraries at  
various stages of development. For a detailed description  
of the library from which this clone was derived, please  
visit our web site at [brainest.eng.uiowa.edu](http://brainest.eng.uiowa.edu).  
TAG\_LIB=NIH\_BMAP\_Ret4\_S2  
TAG\_TISSUE=adult-retina  
TAG\_SEQ=GTACGCGGCAC"

BASE COUNT 73 a 74 c 77 g 74 t

## ORIGIN

Query Match

Best Local Similarity 82.7%; Score 12.4; DB 11; Length 298;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastvtsgwgtt 15

Db 121 TCGAGTTGGAGTT 108

Search completed: February 25, 2002, 17:20:57

Job time: 16150 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:00:08 ; Search time 2331.3 Seconds  
(without alignments)  
176.910 Million cell updates/sec

Title: US-09-698-903B-5

Perfect score: 25

Sequence: 1 ggatcccccgatgagtaagctagc 25

Scoring table: IDENTITY\_NUC

Gapop 10\_0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_om.\*  
20: em\_or.\*  
21: em\_ov.\*  
22: em\_pat.\*  
23: em\_ph.\*  
24: em\_pl.\*  
25: em\_ro.\*  
26: em\_sts.\*  
27: em\_sy.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htgo\_hum.\*  
31: em\_htgo\_inv.\*  
32: em\_htgo\_rod.\*  
33: em\_htg\_hum.\*  
34: em\_htg\_inv.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	6	AX127752 Sequence
2	25	100.0	25	6	AX172446 Sequence
3	25	100.0	249	12	ARGMTUB
4	25	100.0	1037	6	AX10942
5	25	100.0	1085	6	AX10939
6	25	100.0	1160	6	AX10943
c 7	25	100.0	2476	12	TBI251013
c 8	25	100.0	3236	12	TBI251014
c 9	25	100.0	4832	6	AX172441
c 10	25	100.0	4946	6	AX10948
c 11	25	100.0	4946	6	A76915
c 12	25	100.0	4946	6	AR098307
c 13	25	100.0	4945	6	AX172440
c 14	25	100.0	5349	6	A71437
c 15	25	100.0	5560	6	A60112
c 16	25	100.0	5560	6	AR098311
c 17	25	100.0	5865	6	AX127748
c 18	25	100.0	5865	6	AX127748
c 19	25	100.0	6539	6	E31991
c 20	25	100.0	6548	6	A60109
c 21	25	100.0	6548	6	A76916
c 22	25	100.0	6548	6	AR098308
c 23	25	100.0	6548	6	E31990
c 24	25	100.0	7599	6	AX063413
c 25	25	100.0	12095	12	BINHGDNA
c 26	24	96.0	1166	6	AX10941
c 27	24	96.0	3200	6	I44104
c 28	24	96.0	3201	6	I44103
c 29	24	96.0	7566	6	A24783
c 30	24	96.0	7566	6	AR074388
c 31	24	96.0	7639	6	A24782
c 32	24	96.0	7639	6	AR074387
c 33	24	96.0	7811	6	AR078675
c 34	19.2	76.8	134832	2	AF004068
c 35	18.8	75.2	152883	8	AC084319
c 36	18.6	74.4	175826	2	AC021890
c 37	17.8	71.2	25514	2	AC020475
c 38	17.8	71.2	175118	3	AC010842
c 39	17.8	71.2	188272	3	AC005639
c 40	17.8	71.2	298616	3	AE003461
c 41	17.6	70.4	83078	8	AF072897
c 42	17.6	70.4	93489	8	AC005359
c 43	17.6	70.4	149172	9	AC004826
c 44	17.6	70.4	149930	9	AL136084
c 45	17.6	70.4	160570	2	AC092317

ALIGNMENTS

RESULT 1	AX127752	Sequence 5 from Patent WO0131042.	25 bp	DNA	PAT	15-MAY-2001
LOCUS	AX127752	Sequence 5 from Patent WO0131042.				
DEFINITION	AX127752	Sequence 5 from Patent WO0131042.				
ACCESSION	AX127752	Sequence 5 from Patent WO0131042.				
VERSION	AX127752.1	GI:14134399				
KEYWORDS		synthetic construct.				
SOURCE		synthetic construct				
ORGANISM		artificial sequence.				
REFERENCE		1 (bases 1 to 25)				
AUTHORS		Weston,B. and de Beuckeleer,M.				
TITLE		Male-sterile brassica plants and methods for producing same				
JOURNAL		Patent: WO 0131042-A 5 03-MAY-2001;				
FEATURES		Avantis CropScience N.V. (BE)				
source		Location/Qualifiers				
		1..25				
		/organism="synthetic construct"				
		/db_xref="taxon:32630"				
		/note="primer M8B251"				

```
BASE COUNT      6 a      8 c      7 g      4 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatccccgatgagctaagctagc 25
|||||
Db 1 GGATCCCCGATGAGCTAAGCTAGC 25

RESULT 2
AXI72446
LOCUS      AXI72446      25 bp      DNA
DEFINITION Sequence 7 from Patent WO0141558.
ACCESSION AXI72446
VERSION    AXI72446.1 GI:14597558
KEYWORDS   synthetic construct.
SOURCE     synthetic construct
ORGANISM   artificial sequence.
REFERENCE  1 (bases 1 to 25)
AUTHORS    de Both,G. and de Beuckeleer,M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL    Patent: WO 0141558-A 7 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source
            1..25
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="primer 251"

BASE COUNT      6 a      8 c      7 g      4 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatccccgatgagctaagctagc 25
|||||
Db 1 GGATCCCCGATGAGCTAAGCTAGC 25

RESULT 3
ARGMTUB
LOCUS      ARGMTUB      249 bp      DNA
DEFINITION Soybean beta-1-tubulin gene fused to Ti plasmid unit 7 30TR.
ACCESSION X05579
VERSION    X05579.1 GI:58087
KEYWORDS   beta-tubulin; fusion gene; plasmid.
SOURCE     synthetic construct.
ORGANISM   synthetic construct
REFERENCE  1 (bases 1 to 249)
AUTHORS    Guiltinan,M.J., Velten,J., Bustos,M.M., Cyr,R.J., Schell,J. and
Fosket,D.E.
TITLE      The expression of a chimeric soybean beta-tubulin gene in tobacco
JOURNAL    Mol. Gen. Genet. 207, 328-334 (1987)
FEATURES   Location/Qualifiers
            source
            1..249
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="fusion product (17AA); Protein sequence is in
            conflict with the conceptual translation"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA29084.1"
            /db_xref="GI:4376141"
            /translation="AMAAWASSNWSNDPPMS"

CDS
            1..51
            /note="beta-1-tubulin sequence"
            /note="pUC 13 polylinker"
            /note="Sal I linker"
            /note="theroretical fusion junction (24) with gene 7 of Ti
            plasmid"
            173..178
            /note="put.polyA signal"
            198
            /note="polyA site"
            220..225
            /note="put.polyA signal"
            88 t

BASE COUNT      76 a      47 c      38 g      88 t
ORIGIN

Query Match      100.0%; Score 25; DB 12; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatccccgatgagctaagctagc 25
|||||
Db 36 GGATCCCCGATGAGCTAAGCTAGC 60

RESULT 4
AL0942
LOCUS      AL0942      1037 bp      DNA
DEFINITION Nucleotide sequence 4 from patent number DE3920034.
ACCESSION AL0942
VERSION    AL0942.1 GI:492369
KEYWORDS   unidentified.
SOURCE     unidentified
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 1037)
AUTHORS
JOURNAL    Patent: DE 3920034-A 4 31-MAY-1990;
FEATURES   Location/Qualifiers
            source
            1..1037
            /organism="unidentified"
            /db_xref="taxon:32644"

BASE COUNT      338 a      174 c      166 g      359 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 1037;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatccccgatgagctaagctagc 25
|||||
Db 767 GGATCCCCGATGAGCTAAGCTAGC 791

RESULT 5
AL0939
LOCUS      AL0939      1085 bp      DNA
DEFINITION Nucleotide sequence 1 from patent number DE3920034.
ACCESSION AL0939
VERSION    AL0939.1 GI:492367
KEYWORDS   unidentified.
SOURCE     unidentified
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 1085)
AUTHORS
JOURNAL    Patent: DE 3920034-A 1 31-MAY-1990;
FEATURES   Location/Qualifiers
            source
            1..1085
```

```
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      369 a   218 c   155 g   343 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 1085;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaagctagc 25
    |||||||||||||||||||||||||||
Db 815 GGATCCCCCGATGAGCTAAGCTAGC 839

RESULT 6
LOCUS      A10943      1160 bp      DNA      PAT      27-SEP-1993
DEFINITION Nucleotide sequence 5 from patent number DE3920034.
ACCESSION  A10943
VERSION     A10943.1 GI:492370
KEYWORDS   .
SOURCE     .
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 1160)
AUTHORS   .
JOURNAL   Patent: DE 3920034-A 5 31-MAY-1990;
          Location/Qualifiers
FEATURES   source
          1..1160
          /organism="unidentified"
          /db_xref="taxon:32644"
BASE COUNT      367 a   194 c   188 g   411 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 1160;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaagctagc 25
    |||||||||||||||||||||||||||
Db 890 GGATCCCCCGATGAGCTAAGCTAGC 914

RESULT 7
TBI251013/c
LOCUS      TBI251013      2476 bp      DNA      SYN      26-JUN-2001
DEFINITION Transformation binary vector pBAR-A, T-DNA region.
ACCESSION  AJ251013
VERSION     AJ251013.1 GI:6453666
KEYWORDS   t-dna; transformation binary vector.
SOURCE     Transformation binary vector pBAR-A.
ORGANISM   artificial sequence; vectors.
REFERENCE  1 (bases 1 to 2476)
AUTHORS   Cardon,G.H. and Huijser,P.
TITLE     T-DNA sequence of the plant transformation binary vector pBAR-A
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 2476)
AUTHORS   Cardon,G.H.
TITLE     Direct Submission
JOURNAL   Submitted (12-NOV-1999) Cardon G.H., Molecular Plant Genetics,
          Max-Planck- Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg
          10, 50829 Cologne, GERMANY
COMMENT    Related sequences: U09365, J01825, J01636, X02513, V00090, X05822,
          X01077, J01826.
FEATURES   Location/Qualifiers
          source
          1..2476
          /organism="Transformation binary vector pBAR-A"
          /db_xref="taxon:108140"
          /note="can be cultured in Escherichia coli or
          Agrobacterium tumefaciens-T-DNA region of a binary vector

derived from plasmid pgPTV-BAR"
1..148
/note="similar to T-DNA left border"
complement(149..279)
/gene="laci"
CDS
complement(<149..>279)
/gene="laci"
/codon_start=1
/transl_table=11
/protein_id="CAC42834.1"
/db_xref="GI:14572627"
/translation="NHHQTGFSPAGANQRGPIAATLSCPGGEGQSAAVARLTGCKKNHP"
rep_origin
280..576
/note="M13 ori"
/direction=RIGHT
complement(<577..>732)
/gene="lacZ"
gene
complement(577..732)
/gene="lacZ"
733..955
terminator
/gene="transcript 7"
733..955
gene
/gene="transcript 7"
complement(<956..>1544)
/gene="bar"
gene
complement(956..1544)
/gene="bar"
1545..1837
promoter
/gene="NOS"
1545..2314
gene
/gene="NOS"
1838..2157
misc_feature
/gene="NOS"
/note="similar to transposable element Tn5"
2158..2191
misc_feature
/gene="NOS"
2192..2314
promoter
/note="synthetic multiple cloning site"
/gene="NOS"
2315..2476
misc_feature
/note="similar to T-DNA right border"
BASE COUNT      563 a   614 c   699 g   600 t
ORIGIN

Query Match      100.0%; Score 25; DB 12; Length 2476;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaagctagc 25
    |||||||||||||||||||||||||||
Db 961 GGATCCCCCGATGAGCTAAGCTAGC 937

RESULT 8
TBI251014/c
LOCUS      TBI251014      3236 bp      DNA      SYN      26-JUN-2001
DEFINITION Transformation binary vector pBAR-35S, T-DNA region.
ACCESSION  AJ251014
VERSION     AJ251014.1 GI:6453667
KEYWORDS   t-dna; transformation binary vector.
SOURCE     Transformation binary vector pBAR-35S.
ORGANISM   Transformation binary vector pBAR-35S.
REFERENCE  1 (bases 1 to 3236)
AUTHORS   Cardon,G.H. and Huijser,P.
TITLE     T-DNA sequence of the plant transformation binary vector pBAR-A
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 3236)
AUTHORS   Cardon,G.H.
TITLE     Direct Submission
JOURNAL   Submitted (12-NOV-1999) Cardon G.H., Molecular Plant Genetics,
```

Max-Planck- Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg  
10, 50829 Cologne, GERMANY  
Related sequences: U09365, J01825, J01636, X02513, V00090, X05822  
X01077, J01826.  
COMMENT

```

FEATURES             Location/Qualifiers
     1..3236
         /organism="Transformation binary vector pBAR-35S"
         /db_xref="taxon:108141"
         /note="Can be cultured in Escherichia coli or
Agrobacterium tumefaciens-T-DNA region of a binary vector
derived from plasmid pGPtv-BAR"
     1..148
         /misc_feature="similar to T-DNA left border"
         /complement(149..279)
         /gene="laci"
         /complement(<149..>279)
         /gene="laci"
         /codon_start=1
         /transl_table=11
         /protein_id="CAC42835.1"
         /db_xref="GI:14572628"
         /translation="NHQGTGFGPAGANQRPLAATLSGGEGQSAVARLTGKKNHPE"

```

```

misc_feature      280..576
                  /note="M13 ori"
mrna              complement(<577..>732)
                  /gene="lacZ"
gene             complement(577..732)
                  /gene="lacZ"
                  733..955
terminator       /gene="transcript 7"
                  733..955
                  /gene="transcript 7"
                  /note="putative"
mrna             complement(<956..>1544)
                  /gene="bar"
gene            complement(956..1544)
                  1545..1837
promoter        /gene="bar"
                  1545..1837
gene           /gene="NOS"
                  1545..1837
                  /gene="NOS"
misc_feature    1838..2157
                  /note="similar to transposable element Tn5"
promoter       2167..2707
                  /gene="camV 35S"
gene          2167..2956
                  /gene="camV 35S"
misc_feature   2708..2730
                  /gene="camV 35S"
                  /note="Synthetic multiple cloning site"
terminator    2731..2956
                  /gene="camV 35S"
gene         2957..3074
                  /gene="NOS gene promoter"
promoter     2957..3074
                  /gene="NOS gene promoter"
misc_feature  3075..3236
                  /note="similar to T-DNA right border"
BASE COUNT      804 a      786 c      860 g      786 t
ORIGIN

```

Query Match	100.0%	Score 25;	DB 12;	Length 3236;
Best Local Similarity	100.0%;	Pred. No. 0.036;		
Matches 25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ggatcccccgatgactaagctagc	25	
Db	961	GGATCCCCCGATGAGCTAAGCTAGC	937	

**RESULT 9**

```

AX172441/c
LOCUS       AX172441      4832 bp      DNA
DEFINITION  Sequence 2 from Patent WO0141558.
ACCESSION   AX172441
VERSION     AX172441.1  GI:14597553
KEYWORDS    .
SOURCE      synthetic construct.
            ORGANISM      artificial sequence.
            REFERENCE     1 (bases 1 to 4832)
            AUTHORS       de Both,G. and de Beuckeleer,M.
            TITLE         Hybrid winter oilseed rape and methods for producing same
            JOURNAL        Patent: WO 0141558-A 2 14-JUN-2001;
            FEATURES       Aventis CropScience N.V. (BE)
                           location/Qualifiers
                           1..4832
                           /organism="synthetic construct"
                           /db_xref="taxon:32630"
                           /note="T-DNA of plasmid pTHW118"
                           misc_feature      1883..4065
                           /note="HpaI restriction fragment"
                           BASE COUNT      1528 a 883 c 932 g 1488 t 1 others
                           ORIGIN

```

```
Query Match 100.0%; Score 25; DB 6; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 ggatccccgatgagctaagctagc 25  
|||||  
Db 317 GGATCCCCGATGAGCTAAGCTAGC 293

RESULT	10
A60108/c	
LOCUS	A60108
DEFINITION	Sequence 1 from Patent WO9706267.
ACCESSION	A60108
VERSION	A60108.1 GI:3715124
KEYWORDS	
SOURCE	Transformation vector pTHW107.
ORGANISM	Transformation vector pTHW107 artificial sequence; vectors. 1 (bases 1 to 4946) De,B.M.
REFERENCE	Genetic transformation using a PARP inhibitor
AUTHORS	Patent: WO 9706267-A 1 20-FEB-1997;
TITLE	PLANT GENETIC SYSTEMS NV (BE)
JOURNAL	Location/Qualifiers 1..4946
FEATURES	/organism="Transformation vector pTHW107" /db_xref="taxon:126810"
source	
BASE COUNT	1569 a 891 c 963 g 1523 t
ORIGIN	

```

Query Match          100.0%; Score 25; DB 6; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ggatcccccgatgagctaagctagc 25  
|||  
Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT	11	
A76915/c		
LOCUS	A76915	4946 bp
DEFINITION	Sequence 1	DNA
ACCESSION	A76915	from Patent EP0757102.
VERSION	A76915.1	
KEYWORDS		GI:6088712
		19-OCT-1999

```

SOURCE      Transformation vector pTHW107.
ORGANISM    Transformation vector pTHW107
            artificial sequence; vectors.
REFERENCE   1 (bases 1 to 4946)
AUTHORS    De,B.M.
TITLE      GENETIC TRANSFORMATION USING A PARP INHIBITOR
JOURNAL    Patent: EP 0757102-A 1 05-FEB-1997;
            PLANT GENETIC SYSTEMS NV (BE)
FEATURES   Location/Qualifiers
            source
            1..4946
            /organism="Transformation vector pTHW107"
            /db_xref="taxon:126810"
BASE COUNT  1569 a 891 c 963 g 1523 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25
    |||||||||||||||||||||||||||
Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT 12
AR098307/c      AR098307      4946 bp      DNA      14-FEB-2001
LOCUS           Sequence 1 from patent US 6074876.
ACCESSION       AR098307
VERSION         AR098307.1 GI:12807564
KEYWORDS        .
SOURCE          Unknown.
ORGANISM        Unclassified.
REFERENCE       1 (bases 1 to 4946)
AUTHORS        De Block,M.
TITLE          Genetic transformation using a PARP inhibitor
JOURNAL        Patent: US 6074876-A 1 13-JUN-2000;
FEATURES       Location/Qualifiers
            source
            1..4946
            /organism="unknown"
BASE COUNT     1569 a 891 c 963 g 1523 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25
    |||||||||||||||||||||||||||
Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT 13
AX172440/c      AX172440      4946 bp      DNA      03-JUL-2001
LOCUS           Sequence 1 from Patent WO0141558.
ACCESSION       AX172440
VERSION         AX172440.1 GI:14597552
KEYWORDS        .
SOURCE          synthetic construct.
ORGANISM        synthetic construct.
REFERENCE       1 (bases 1 to 4946)
AUTHORS        de Both,G. and de Beuckeleer,M.
TITLE          Hybrid winter oilseed rape and methods for producing same
JOURNAL        Patent: WO 0141558-A 1 14-JUN-2001;
            Aventis CropScience N.V. (BE)
FEATURES       Location/Qualifiers
            source
            1..4946
            /organism="synthetic construct"

SOURCE      Transformation vector pTHW107.
ORGANISM    Transformation vector pTHW107
            artificial sequence; vectors.
REFERENCE   1 (bases 1 to 4946)
AUTHORS    De,B.M.
TITLE      GENETIC TRANSFORMATION USING A PARP INHIBITOR
JOURNAL    Patent: EP 0757102-A 1 05-FEB-1997;
            PLANT GENETIC SYSTEMS NV (BE)
FEATURES   Location/Qualifiers
            source
            1..4946
            /organism="Transformation vector pTHW107"
            /db_xref="taxon:126810"
BASE COUNT  1569 a 891 c 963 g 1523 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25
    |||||||||||||||||||||||||||
Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT 14
A71437/c      A71437      5349 bp      DNA      07-MAY-1999
LOCUS           Sequence 7 from Patent WO9810081.
ACCESSION       A71437
VERSION         A71437.1 GI:4775050
KEYWORDS        .
SOURCE          unidentified.
ORGANISM        unidentified.
REFERENCE       1 (bases 1 to 5349)
AUTHORS        Michiels,F. and Williams,M.
TITLE          IMPROVED BARSTAR GENE
JOURNAL        Patent: WO 9810081-A 7 12-MAR-1998;
            MICHELIS FRANK (BE)
FEATURES       Location/Qualifiers
            source
            1..5349
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT     1339 a 1233 c 1290 g 1487 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5349;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25
    |||||||||||||||||||||||||||
Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT 15
A60112/c      A60112      5560 bp      DNA      06-MAR-1998
LOCUS           Sequence 5 from Patent WO9706267.
ACCESSION       A60112
VERSION         A60112.1 GI:3715128
KEYWORDS        .
SOURCE          Transformation vector pTHW142.
ORGANISM        Transformation vector pTHW142.
REFERENCE       1 (bases 1 to 5560)
AUTHORS        De,B.M.
TITLE          Genetic transformation using a PARP inhibitor
JOURNAL        Patent: WO 9706267-A 5 20-FEB-1997;
            PLANT GENETIC SYSTEMS NV (BE)
FEATURES       Location/Qualifiers
            source
            1..5560
            /organism="Transformation vector pTHW142"
            /db_xref="taxon:126817"
BASE COUNT     1533 a 1199 c 1244 g 1576 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 5560;
```

Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ggatcccccgatgagctaagctagc 25  
|||||  
Db 304 GGATCCCCCGAGGAGCTAGC 280

Search completed: February 25, 2002, 18:00:10  
Job time: 18418 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:22 ; Search time 716.55 Seconds  
(without alignments)  
29.912 Million cell updates/sec

Title: US-09-698-903B-5

Perfect score: 25

Sequence: 1 ggatcccccatgagctaagctagc 25

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N\_Geneseq\_1101.\*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*
  - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*
  - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*
  - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*
  - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*
  - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*
  - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*
  - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*
  - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*
  - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*
  - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*
  - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*
  - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*
  - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*
  - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*
  - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*
  - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*
  - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*
  - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*
  - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*
  - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*
  - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	25	100.0	25	22	AAH25426	PCR primer for fla
2	25	100.0	25	22	AAH25426	PCR primer MDB251
3	25	100.0	1037	11	AAQ04705	USP-Promoter-casse
4	25	100.0	1085	11	AAQ04703	Legumin-signalpept
5	25	100.0	1160	11	AAQ04706	USP-signalpeptide
6	25	100.0	1303	17	AAAT39337	Plasmid pTS88 (Eco
7	25	100.0	4832	22	AAH25423	Nucleotide sequenc
8	25	100.0	4946	18	AAH25431	T-DNA of plasmid p
9	25	100.0	4946	22	AAH25422	Nucleotide sequenc
10	25	100.0	5228	22	AAH25439	Plasmid pTS172delt
11	25	100.0	5349	19	AAV23239	T-DNA of pTTS24.

12	25	100.0	5864	17	AAT39339	Plasmid pTColl3 T-
13	25	100.0	5864	17	AAT39339	Chimeric T-DNA of
14	25	100.0	5865	22	AAH25426	Chimeric T-DNA of
15	25	100.0	5865	22	AAH25426	E. coli plasmid pT
16	25	100.0	6539	21	AAZ91097	Plasmid pTS174 use
17	25	100.0	6548	17	AAT39336	Plasmid pTS172. C
18	25	100.0	6548	18	AAT61394	E. coli plasmid pT
19	25	100.0	6548	21	AAZ91096	Plasmid pTS346. U
20	25	100.0	7492	22	AAH25426	Nucleotide sequenc
21	25	100.0	7599	22	AAH25426	USP-Promoter-casse
22	24	96.0	1166	11	AAQ04704	USP-Promoter-casse
23	24	96.0	3201	12	AAQ14529	USP-Promoter-casse
24	24	96.0	3201	12	AAQ14529	USP-Promoter-casse
25	24	96.0	7566	14	AAQ42160	Plasmid pTS346. U
26	24	96.0	7639	14	AAQ42160	Plasmid pTS346. U
27	19	76.0	34	22	AAH25165	PCR primer used to
28	19	76.0	34	22	AAH25165	Human polynucleoti
29	17.2	68.8	2562	22	AAI59154	Human polynucleoti
30	17.2	68.8	6380	22	AAI59154	Human polynucleoti
31	17.2	68.8	7104	22	AAI59608	Human polynucleoti
32	17	68.0	29	14	AAQ42148	Plasmid pTS174 use
33	17	68.0	3153	21	AAZ29122	Plasmid pTS174 use
34	17	68.0	3336	21	AAZ29121	Plasmid pTS174 use
35	17	68.0	3694	21	AAZ29121	Plasmid pTS174 use
36	17	68.0	3877	21	AAZ29123	Plasmid pTS174 use
37	17	68.0	24593	6	AAH250226	Sequence of opine
38	17	68.0	24596	6	AAH250226	Complete nucleotid
39	16.6	66.4	3488	19	AAZ96307	S. pneumoniae deri
40	16.6	66.4	6204	19	AAV44592	Streptococcus pneu
41	16.6	66.4	534720	19	AAV30458	Rhizobium species
42	16.6	66.4	536165	19	AAV30459	Rhizobium species
43	16.2	64.8	457	21	AAH25426	Arabidopsis thalia
44	16.2	64.8	521	21	AAH25426	Arabidopsis thalia
45	16.2	64.8	526	21	AAH25426	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAH25426 standard; DNA; 25 BP.

AC AAH25426;

DT 22-AUG-2001 (first entry)

XX PCR primer for flanking regions in transgenic plant MS-BN1.

XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
KW fertility restorer gene; barnase gene; PCR primer; ss.

XX Synthetic.

XX WO200141558-A1.

PD 14-JUN-2001.

PF 06-DEC-2000; 2000WO-EP12872.

XX 08-DEC-1999; 99US-0457037.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX De Both G, De Beuckeleer M;

XX WPI; 2001-381419/40.

PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
PT with improved qualities, comprises a male-sterility gene and fertility  
XX restorer gene, integrated into the genome -  
PS Example 4; Page 42; 98pp; English.

XX CC The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAH25425-26 were used to amplify the flanking region of  
 CC a vector in a transgenic plant which carries the TA29-barnase  
 CC transgene.  
 XX SQ Sequence 25 BP; 6 A; 8 C; 7 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0053;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25  
 |||||  
 Db 1 ggatcccccgatgagctaaagctagc 25

RESULT 2  
 AAD06994  
 ID AAD06994 standard; DNA; 25 BP.  
 AC AAD06994;  
 DT 06-AUG-2001 (first entry)  
 DE PCR primer MDB251 to generate the flanking region of elite event MS-B2.  
 KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; thermal asymmetric interlaced;  
 KW TAIL; ss.  
 OS Agrobacterium sp.  
 PN WO200131042-A2.  
 XX 03-MAY-2001.  
 XX 26-OCT-2000; 2000WO-EP10680.  
 XX 29-OCT-1999; 99US-0430497.  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX Weston B, De Beuckeleer M;  
 XX WPI; 2001-300517/31.  
 XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 XX harboring specific transformation events, particularly by presence of  
 XX male-sterility gene, at specific location in its genome -  
 XX Example 3; Page 28; 53pp; English.  
 XX The present invention relates to a transgenic Brassica plant or its  
 XX seed, cells or tissues, characterised by harbouring a specific  
 XX transformation event, particularly by the presence of a male-sterility  
 XX gene, at a specific location in the Brassica genome. Transgenic  
 XX Brassica plant is useful for producing a hybrid seed by crossing the  
 XX transgenic plant with a male-fertile Brassica plant and harvesting the  
 XX hybrid seed from the transgenic Brassica plant.  
 XX The present sequence is primary thermal interlaced (TAIL)-PCR primer  
 XX MDB251 used to right (5') and left (3') border flanking region of elite  
 XX event MS-B2. This primer corresponds to position 293-317 of plasmid  
 XX pTCO113.

XX SQ Sequence 25 BP; 6 A; 8 C; 7 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0053;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25  
 |||||  
 Db 1 ggatcccccgatgagctaaagctagc 25

RESULT 3  
 AAQ04705  
 ID AAQ04705 standard; DNA; 1037 BP.  
 AC AAQ04705;  
 DT 12-OCT-1990 (first entry)  
 XX USP-Promoter-cassette USP-Pr.T7.1.  
 DE Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.  
 XX DE3920034-A.  
 XX 31-MAY-1990.  
 XX 20-JUN-1989; 89DE-3920034.  
 XX 19-SEP-1988; 88DD-0319887.  
 XX (PFLA-) VE KOMB PFLANZENZUC.  
 XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;  
 XX WPI; 1990-172459/23.  
 XX Incorporation of DNA into higher plant genome - by specified  
 XX recombinant DNA techniques.  
 XX Disclosure; ; pp; German.  
 XX The unique BglII-Ort (720-725) site is for ligating foreign DNA and the  
 XX HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 XX cassette in the TI-vector pGA471. Agrobacterium tumefaciens is  
 XX transfected.  
 XX See also AAQ04703-Q04706.  
 XX Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 100.0%; Score 25; DB 11; Length 1037;  
 Best Local Similarity 100.0%; Pred. No. 0.007;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25  
 |||||  
 Db 767 ggatcccccgatgagctaaagctagc 791

RESULT 4  
 AAQ04703  
 ID AAQ04703 standard; DNA; 1085 BP.  
 AC AAQ04703;  
 XX 12-OCT-1990 (first entry)  
 DE Legumin-signalpeptide cassette Le-Sig.T7.  
 XX



```
KW Foreign DNA incorporation; recombinant DNA techniques;
XX higher plant genome; legumin; B4 gene; signalpeptide; Le-Sig.T7.; ss.
FH Key Location/Qualifiers
FT CDS 747..814
FT /*tag= a
FT /product=Legumin-signalpeptide
XX
XX
XX DE3920034-A.
XX
XX 31-MAY-1990.
XX
XX 20-JUN-1989; 89DE-3920034.
XX
XX 19-SEP-1988; 88DD-0319887.
XX
XX (PFLA-) VE KOMB PFLANZENZUC.
XX
XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;
XX
XX WPI; 1990-172459/23.
XX P-PSDB; AAR05198.
XX
XX Incorporation of DNA into higher plant genome - by specified
XX recombinant DNA techniques.
XX
XX Disclosure; ; pp; German.
XX
XX The Legumin gene B4 is used. The unique BglII-Ort (815-820) site is for
XX ligating foreign DNA and the HindIII-Ort in the 3' polylinker
XX (1080-1085) for cloning the cassette in the Ti-vector pGA471.
XX The cassette is cloned into the binary Ti-vectors pGA471 and
XX Agrobacterium tumefaciens is transfected.
XX See also AAQ04703-Q04706.
XX
XX Sequence 1085 BP; 368 A; 219 C; 155 G; 343 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 25; DB 11; Length 1085;
XX Best Local Similarity 100.0%; Pred. No. 0.007;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 ggatcccccgatgagctaaagctagc 25
XX |||||||||||||||||||
XX Db 815 ggatcccccgatgagctaaagctagc 839
XX
XX RESULT 5
XX AAQ04706
XX ID AAQ04706 standard; DNA; 1160 BP.
XX
XX AC AAQ04706;
XX
XX DT 12-OCT-1990 (first entry)
XX
XX USP-signalpeptide cassette USP-sig.T7.
XX
XX Foreign DNA incorporation; recombinant DNA techniques;
XX higher plant genome; signalpeptide; USP-Sig.T7.; ss.
XX
XX Key Location/Qualifiers
XX FH 708..877
XX FT CDS /*tag= a
XX FT /product=signalpeptide
XX FT 747..817
XX
XX intron
XX
XX DE3920034-A.
XX
XX 31-MAY-1990.
XX
XX 20-JUN-1989; 89DE-3920034.
XX
XX 19-SEP-1988; 88DD-0319887.
XX
```

```
XX (PFLA-) VE KOMB PFLANZENZUC.
XX
XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;
XX
XX WPI; 1990-172459/23.
XX P-PSDB; AAR05199.
XX
XX Incorporation of DNA into higher plant genome - by specified
XX recombinant DNA techniques.
XX
XX Disclosure; ; pp; German.
XX
XX The unique BglII-Ort (890-895) site is for
XX ligating foreign DNA and the HindIII-Ort in the 3' polylinker
XX (1155-1160) for cloning the cassette in the Ti-vector pGA471.
XX The cassette is cloned into the binary Ti-vectors pGA471 and
XX Agrobacterium tumefaciens is transfected.
XX See also AAQ04703-Q04706.
XX
XX Sequence 1160 BP; 369 A; 192 C; 188 G; 411 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 25; DB 11; Length 1160;
XX Best Local Similarity 100.0%; Pred. No. 0.007;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 ggatcccccgatgagctaaagctagc 25
XX |||||||||||||||||||
XX Db 890 ggatcccccgatgagctaaagctagc 914
XX
XX RESULT 6
XX AAT39337
XX ID AAT39337 standard; DNA; 1303 BP.
XX
XX AC AAT39337;
XX
XX DT 22-JAN-1997 (first entry)
XX
XX Plasmid pTS88 (EcoRI-HindIII fragment).
XX
XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
XX transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FH 1..35
XX FT misc_feature /*tag= a
XX FT /label= pGEM2
XX FT /note= "polylinker of pGEM2"
XX FT 36..694
XX FT promoter /*tag= b
XX FT /label= P35S
XX FT /function= 35S promoter of cauliflower mosaic virus
XX FT strain CM1841
XX
XX CDS 695..967
XX FT /*tag= c
XX FT /label= barstar
XX FT /product= Bacillus amyloliquefaciens barstar
XX FT 968..1287
XX FT /tag= d
XX FT /label= 3'g7
XX FT /function= region containing polyadenylation signal
XX FT 1288..1303 of gene 7 og Agrobacterium T-DNA
XX FT /*tag= e
XX FT /label= pGEM2
XX FT /note= "polylinker of pGEM2"
XX
XX PN WO9626283-A1.
XX
```

```

PD 29-AUG-1996.
XX
PF 21-FEB-1996; 96WO-EP00722.
XX
PR 21-FEB-1995; 95EP-0400364.
XX
PA (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
PI Botterman J, Cornelissen M, Michiels F;
XX WPI; 1996-402373/40.
XX
DR Prodn. of male sterile plants by transforming with a chimaeric
PT construct - comprising a male sterility DNA e.g. barnase and a
PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
PT for generating hybrid cultivars
XX
XX Example 1; Page 38; 56pp; English.
XX
CC The HindIII-EcoRI fragment (AAT39337) of plasmid pT888 contains
CC barstar DNA under control of a 35S promoter. The plasmid was
CC used with pT8174 (see also AAT39336) contg. barnase DNA under
CC control of the stamen-specific promoter El to produce male sterile
CC rice cv. Kochihibiki transgenic plants, and with plasmid pVE136
CC (see also AAT39338) transg. barnase DNA under control of the stamen-
CC specific PCA35 promoter to produce male sterile maize plants.
CC Expression of barnase (a ribonuclease) in the stamen leads to male
CC sterility. Constitutive expression of barstar counteracts possible
CC low level expression of barnase DNA in non-stamen tissue.
XX
SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 100.0%; Score 25; DB 17; Length 1303;
Best Local Similarity 100.0%; Pred. No. 0.0071;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaagctagc 25
Db 1024 ggatcccccgatgagctaagctagc 1048

RESULT 7
AAH25423/c
ID AAH25423 standard; DNA; 4832 BP.
XX
AC AAH25423;
XX
XX 22-AUG-2001 (first entry)
XX
DE Nucleotide sequence of plasmid pTHW118.
XX
KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barstar gene; ss.
XX
OS Synthetic.
OS Streptomyces hygroscopicus.
OS Arabidopsis thaliana.
OS Bacillus amyloliquefaciens.
OS Nicotiana tabacum.
XX
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /*tag= a
FT /*note= "right border repeat from TL-DNA from pTiB6S3"
FT misc_feature 26..53
FT /*tag= b
FT /*note= "synthetic polylinker derived sequences"
FT misc_feature 54..90
FT /*tag= c
FT /*note= "residual sequence from TL-DNA at right
FT misc_feature 91..97
FT border repeat"

```

```

FT FT /*tag= d
FT FT /*note= "synthetic polylinker derived sequences"
FT FT complement (98..309)
FT FT /*tag= e
FT FT /*note= "3' UTR from TL-DNA gene 7 of pTiB6S3"
FT FT misc_feature 310..330
FT FT /*tag= f
FT FT /*note= "synthetic polylinker derived sequences"
FT FT complement (331..882)
FT FT /*tag= g
FT FT /*note= "Streptomyces hygroscopicus bialaphos
FT FT resistance (bar) gene"
FT FT complement (883..2608)
FT FT /*tag= h
FT FT /*note= "atsIA ribulose-1,5-biphosphate carboxylase
FT FT small subunit gene from Arabidopsis thaliana"
FT FT misc_feature 2609..2658
FT FT /*tag= i
FT FT /*note= "synthetic polylinker derived sequences"
FT FT complement (2659..2919)
FT FT /*tag= j
FT FT /*note= "TaqI fragment from 3' UTR of nopaline
FT FT synthase gene from T-DNA of pTiT37 and
FT FT containing plant polyadenylation signals"
FT FT misc_feature 2920..2940
FT FT /*tag= k
FT FT /*note= "synthetic polylinker derived sequences"
FT FT 2941..2980
FT FT /*tag= l
FT FT /*note= "downstream of Bacillus amyloliquefaciens
FT FT barstar coding region"
FT FT complement (2981..3253)
FT FT /*tag= m
FT FT /*note= "Barstar gene coding region from Bacillus
FT FT amyloliquefaciens"
FT FT complement (3254..4762)
FT FT /*tag= n
FT FT /*note= "anther-specific gene TA29 promoter from
FT FT Nicotiana tabacum"
FT FT misc_feature 4763..4807
FT FT /*tag= o
FT FT /*note= "synthetic polylinker derived sequences"
FT FT 4808..4832
FT FT /*tag= p
FT FT /*note= "left border repeat from TL-DNA from pTiB6S3"
XX
PN WO200141558-A1.
XX
XX 14-JUN-2001.
PD
XX
XX 06-DEC-2000; 2000WO-EP12872.
PF
XX
XX 08-DEC-1999; 99US-0457037.
PR
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
PA
XX
XX De Both G, De Beuckeleer M;
PI
XX WPI; 2001-381419/40.
XX
XX Transgenic winter oilseed rape plants suited for producing hybrid seed
PT with improved qualities, comprises a male-sterility gene and fertility
PT restorer gene, integrated into the genome -
XX
XX Example 1; Page 80-82; 98pp; English.
XX
CC The specification describes a pair of transgenic winter oilseed rape
CC plants suited for producing hybrid seed. One of the plants has an
CC expression cassette comprising a male-sterility gene, and the other
CC plant has an expression cassette comprising a fertility restorer gene,
CC integrated into the genome. The fertility restorer gene is capable of
CC preventing the activity of the male-sterility gene. The plant pair is
CC useful for producing hybrid seed. Plants developed from the hybrid
CC

```

CC seed have agronomic performance, genetic stability and adaptability to  
 CC different genetic backgrounds. The present sequence represents  
 CC plasmid pTHW118. This plasmid comprises the barstar gene, which acts as  
 CC a fertility restorer gene. The plasmid is used to create transgenic  
 CC plants of the invention.

XX SQ Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;

Query Match 100.0%; Score 25; DB 22; Length 4832;  
 Best Local Similarity 100.0%; Pred. No. 0.0078;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaagctagc 25  
 |||||  
 Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT 8  
 AAT59531/c  
 ID AAT59531 standard; DNA; 4946 BP.

AC AAT59531;

XX 07-MAY-1997 (first entry)

DE T-DNA of plasmid pTHW107.

XX Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PAPP;  
 KW niacinamide; Agrobacterium; T-DNA; male sterile; barnase;  
 KW ribonuclease; Rns; cereal; wheat; oilseed rape; Brassica napus;  
 KW plasmid pTHW107; ds.

XX Chimeric Agrobacterium sp.;  
 OS Chimeric Arabidopsis thaliana;  
 OS Chimeric Nicotiana tabacum.

FH Key Location/Qualifiers  
 FT misc\_RNA complement (1..25)

FT /tag= a  
 FT /label= RB  
 FT /note= "T-DNA right border"  
 FT complement (97..330)  
 FT /tag= b  
 FT /label= 3'g7  
 FT /note= "3' untranslated region contg. the poly-A  
 FT signal of gene-7 of Agrobacterium T-DNA"

FT CDS complement (331..882)  
 FT /tag= c  
 FT /label= Bar  
 FT /product= phosphinothricin acetyltransferase  
 FT complement (883..2608)  
 FT /tag= d  
 FT /label= PSSU  
 FT /note= "promoter region of Rubisco small subunit  
 FT gene of Arabidopsis thaliana"

FT promoter complement (2658..3031)  
 FT /tag= e  
 FT /label= 3'nos  
 FT /note= "3' untranslated region contg. the poly-A  
 FT signal of the nopaline-synthase gene of  
 FT Agrobacterium T-DNA"

FT CDS complement (3032..3367)  
 FT /tag= f  
 FT /label= Barnase  
 FT /product= barnase  
 FT complement (3368..4876)  
 FT /tag= g  
 FT /label= PTA29  
 FT /note= "promoter region of tobacco TA29 gene"

FT misc\_RNA complement (4822..4946)  
 FT /tag= h  
 FT /label= LB

FT /note= "T-DNA left border"  
 XX EP757102-A1.  
 PN  
 XX  
 PD 05-FEB-1997.

XX 04-AUG-1995; 95EP-0401844.

XX 04-AUG-1995; 95EP-0401844.

XX (PLBZ ) PLANT GENETIC SYSTEMS NV.

XX De Block M;

XX WPI; 1997-1111050/11.

XX Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase  
 FT inhibitor - reduces the cultured cells response to stress and  
 FT reduces metabolism

XX Example 3; Page 13-16; 25pp; English.

XX Plasmid pTHW107 is a vector carrying T-DNA (AAT59531) comprising a  
 CC barnase coding sequence under control of the tobacco TA29 gene  
 CC stamen-specific promoter and a phosphinothricin acetyltransferase  
 CC coding sequence under control of an Arabidopsis Rubisco small  
 CC subunit gene promoter. Oilseed rape hypocotyl explants were  
 CC infected with Agrobacterium tumefaciens C58C1Rif carrying vector  
 CC pTHW107 and helper Ti plasmid pMP60. In some treatments, the  
 CC hypocotyls were treated with the poly-(ADP-ribose) polymerase  
 CC inhibitor niacinamide (250 mg/l) 4 days prior to infection.  
 CC Plants regenerated from niacinamide-treated transformed calli  
 CC had a low copy number and displayed less variation in the  
 CC expression profile of the transgenes.

XX Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match 100.0%; Score 25; DB 18; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.0078;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaagctagc 25  
 |||||  
 Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT 9  
 AAH25422/c  
 ID AAH25422 standard; DNA; 4946 BP.

XX AC AAH25422;

XX 22-AUG-2001 (first entry)

DE Nucleotide sequence of plasmid pTHW107.

XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; ss.

XX Synthetic.

OS Streptomyces hygroscopicus.

OS Arabidopsis thaliana.

OS Bacillus amyloliquefaciens.

OS Nicotiana tabacum.

XX Key Location/Qualifiers

FT misc\_feature 1..25

FT /tag= a

FT /note= "right border repeat from TL-DNA from pTHB6S3"

FT misc\_feature 26..97

FT /tag= b

FT /note= "synthetic polylinker derived sequences"

FT 3'UTR complement (98..309)  
FT /tag= c  
FT /note= "3' UTR from TL-DNA gene 7 of pT1B6S3"  
FT 310..330  
FT /tag= d  
FT /note= "synthetic polylinker derived sequences"  
FT 331..882  
FT /tag= e  
FT /note= "Streptomyces hygrosopicus bar gene"  
FT complement (883..2608)  
FT /tag= f  
FT /note= "ats1A ribulose-1,5-biphosphate carboxylase  
FT small subunit gene from Arabidopsis thaliana"  
FT 2609..2658  
FT /tag= g  
FT /note= "synthetic polylinker derived sequences"  
FT complement (2659..2919)  
FT /tag= h  
FT /note= "taqI fragment from 3' UTR of nopaline  
FT synthase gene from T-DNA of pT137 and  
FT containing plant polyadenylation signals"  
FT 2920..3031  
FT /tag= i  
FT /note= "3'UTR downstream of Bacillus amyloliquefaciens  
FT barnase coding region"  
FT complement (3032..3367)  
FT /tag= j  
FT /note= "Barnase coding region from Bacillus  
FT amyloliquefaciens"  
FT complement (3368..4877)  
FT /tag= k  
FT /note= "anther-specific gene TA29 promoter from  
FT Nicotiana tabacum"  
FT 4878..4921  
FT /tag= l  
FT /note= "synthetic polylinker derived sequences"  
FT 4922..4946  
FT /tag= m  
FT /note= "left border repeat from TL-DNA from pT1B6S3"  
FT WO200141558-A1.  
XX  
PD 14-JUN-2001.  
XX  
PF 06-DEC-2000; 2000WO-EPI2872.  
XX  
PR 08-DEC-1999; 99US-0457037.  
XX  
PA (AVET ) AVENTIS CROPSCEINCE NV.  
XX  
PI De Both G, De Beuckeleer M;  
XX  
DR WPI; 2001-381419/40.  
XX  
PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
PT with improved qualities, comprises a male-sterility gene and fertility  
PT restorer gene, integrated into the genome -  
XX  
PS Example 1; Page 78-80; 98pp; English.  
XX  
CC The specification describes a pair of transgenic winter oilseed rape  
CC plants suited for producing hybrid seed. One of the plants has an  
CC expression cassette comprising a male-sterility gene, and the other  
CC plant has an expression cassette comprising a fertility restorer gene,  
CC integrated into the genome. The fertility restorer gene is capable of  
CC preventing the activity of the male-sterility gene. The plant pair is  
CC useful for producing hybrid seed. Plants developed from the hybrid  
CC seed have agronomic performance, genetic stability and adaptability to  
CC different genetic backgrounds. The present sequence represents  
CC a plasmid pT1W107. This plasmid comprises the barnase gene, which acts as  
CC a male-sterility gene. The plasmid is used to create transgenic plants  
CC of the invention.

SQ Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;  
Query Match 100.0%; Score 25; DB 22; Length 4946;  
Best Local Similarity 100.0%; Pred. No. 0.0078;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ggatcccccgatgagctaaagctagc 25  
Dd 317 ggatcccccgatgagctaaagctagc 293  
RESULT 10  
AAF86439  
ID AAF86439 standard; DNA; 5228 BP.  
XX  
AC AAF86439;  
XX  
XX 25-JUN-2001 (first entry)  
XX Plasmid pT1S172delta.  
XX Male sterile plant; RNAase inhibitor; plasmid pT1S172delta; ds.  
XX Unidentified.  
XX OS  
XX PN WO200124616-A1.  
XX PD 12-APR-2001.  
XX PF 12-SEP-2000; 2000WO-JP06222.  
XX PR 30-SEP-1999; 99JP-0279307.  
XX PA (NISB ) JAPAN TOBACCO INC.  
XX PI Hamada K, Nakakido F;  
XX DR WPI; 2001-266212/27.  
XX  
PT Method for producing male sterile rice and maize by inserting RNAse  
PT gene and RNAse inhibitor genes with promoters into the plant genome -  
XX  
PS Disclosure; Page 14-17; 29pp; Japanese.  
XX  
CC The present invention relates to a method for producing male sterile  
CC plants. The method comprises inserting a promoter fragment upstream of an  
CC RNAse gene and a second promoter, upstream of an RNAse inhibitor protein  
CC gene and inserting it into the plant genome. The method is useful for  
CC producing male sterile tobacco, lettuce and rapeseed plants, but  
CC preferably rice and maize. The present sequence is a vector used in  
CC the method of the present invention.  
XX  
SQ Sequence 5228 BP; 1384 A; 1307 C; 1263 G; 1274 T; 0 other;  
Query Match 100.0%; Score 25; DB 22; Length 5228;  
Best Local Similarity 100.0%; Pred. No. 0.0079;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ggatcccccgatgagctaaagctagc 25  
Dd 4957 ggatcccccgatgagctaaagctagc 4981  
RESULT 11  
AAV23239/C  
ID AAV23239 standard; DNA; 5349 BP.  
XX  
AC AAV23239;  
XX  
DT 17-JUL-1998 (first entry)  
XX

```
DE T-DNA of pTTS24.
XX
KW Barstar; barnase inhibitor; fertility restoration;
KW male-sterile line; plasmid pTTS24; T-DNA; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature complement (1..25)
FT /*tag= a
FT /label= RB
FT /note= "right boarder"
FT /*tag= b
FT /label= 3' g7
FT /note= "region containing 3' untranslated end of
FT Agrobacterium T-DNA gene 7"
FT
FT CDS 332..883
FT /*tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin acetyl
FT transferase"
FT complement (884..2258)
FT /*tag= d
FT /label= P35S
FT /note= "35S promoter of Cauliflower Mosaic Virus"
FT 2281..3969
FT /*tag= e
FT /label= PE1
FT /note= "promoter of E1 gene of rice (W09213956)"
FT 3970..4245
FT /*tag= f
FT /product= improved_barstar
FT 4246..4577
FT /*tag= g
FT /label= 3'_chs
FT /note= "region containing 3' untranslated end of
FT chalcone synthase gene"
FT complement (5325..5349)
FT /*tag= h
FT /note= "T-DNA left border"
XX
PN W09810081-A2.
XX
XX 12-MAR-1998.
XX
XX 01-SEP-1997; 97WO-EP04739.
XX
XX 03-SEP-1996; 96EP-0202446.
XX
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX Michiels F, Williams M;
XX
XX WPI; 1998-193630/17.
XX
XX DNA encoding an improved barstar protein - used to restore fertility
XX in male-sterile plant lines
XX
XX Example 4; Pages 41-43; 54pp; English.
XX
XX The present sequence was used in the preparation of an improved
XX Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which
XX can be used to restore fertility to male-sterile lines.
XX The DNA sequence encoding the improved barstar, leads to increased
XX barstar production in tapetum cells, due to improved translation,
XX and possibly protein stability.
XX
XX Sequence 5349 BP; 1339 A; 1233 C; 1290 G; 1487 T; 0 other;
XX
XX Query Match 100.0%; Score 25; DB 19; Length 5349;
XX Best Local Similarity 100.0%; Pred. No. 0.0079;
```

```
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggatcccccgatgacgactaagctagc 25
   |||||
DB 317 GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT 12
AAT39339
ID AAT39339 standard; DNA; 5864 BP.
XX
AC AAT39339;
XX
DT 22-JAN-1997 (first entry)
XX
DE Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.
XX
KW Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;
KW transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature complement (1..25)
FT /*tag= a
FT /label= RB
FT /note= "right border of Agrobacterium T-DNA"
FT complement (98..330)
FT /*tag= b
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT complement (331..882)
FT /*tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin
FT acetyltransferase"
FT complement (883..2608)
FT /*tag= d
FT /label= PSSu
FT /note= "promoter of Arabidopsis Rubisco small
FT subunit gene"
FT complement (2659..3031)
FT /*tag= e
FT /label= 3'nos
FT /note= "region containing polyA signal of nopaline
FT synthase gene of Agrobacterium T-DNA"
FT complement (3032..3367)
FT /*tag= f
FT /label= Barnase
FT /note= "Bacillus amyloliquefaciens barnase coding
FT region"
FT complement (3368..4877)
FT /*tag= g
FT /label= pTA29
FT /note= "promoter of stamen-specific TA29 gene of
FT Nicotiana tabacum"
FT 4924..5216
FT /*tag= h
FT /label= Pnos
FT /note= "promoter of nopaline synthase gene of
FT Agrobacterium T-DNA"
FT 5217..5489
FT /*tag= i
FT /label= Barstar
FT /note= "region coding for barstar of Bacillus
FT amyloliquefaciens"
FT 5490..5765
FT /*tag= j
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT complement (5840..5864)
FT misc_feature
```

```

FT      /*tag= k
FT      /label= LB
FT      /note= "left border of Agrobacterium T-DNA"
PN      WO9626283-A1.
XX
XX      29-AUG-1996.
XX
XX      21-FEB-1996; 96WO-EP00722.
XX
XX      21-FEB-1995; 95EP-0400364.
XX
XX      (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX      Botterman J, Cornelissen M, Michiels F;
XX      WPI; 1996-402373/40.
XX
XX      Prodn. of male sterile plants by transforming with a chimaeric
XX      construct - comprising a male sterility DNA e.g. barnase and a
XX      co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX      for generating hybrid cultivars
XX
XX      Example 3; Page 33-3743-47; 56pp; English.
XX
XX      Plasmid pTCO113 (AAT39339) is a T-DNA vector containing a bar gene
XX      under control of the PSSU promoter, a barnase gene under control
XX      of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
XX      gene under control of the Pnos promoter. 87% Of oilseed rape
XX      plants regenerated after Agrobacterium-mediated transformation
XX      using pTCO113 were male sterile. Barnase expression disturbed the
XX      function of stamen cells leading to male sterility. Constitutive
XX      expression of barstar counteracted any low level expression of
XX      barnase in non-stamen tissue.
XX
XX      Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;

Query Match      100.0%; Score 25; DB 17; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggatccccgatgagctaagctagc 25
Db      |||||
5546 ggatccccgatgagctaagctagc 5570

RESULT 13
AAT39339/C
ID      AAT39339 standard; DNA; 5864 BP.
XX
XX      AAT39339;
XX
XX      22-JAN-1997 (first entry)
XX
XX      Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.
XX
XX      Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;
XX      transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX      Synthetic.
XX
XX      Key      Location/Qualifiers
XX      misc_feature complement (1..25)
XX      /*tag= a
XX      /label= RB
XX      /note= "right border of Agrobacterium T-DNA"
XX      complement (98...330)
XX      /*tag= b
XX      /label= 3'g7
XX      /note= "region containing polyA signal of gene 7
XX      of Agrobacterium T-DNA"
XX      complement (331..882)
XX      CDS

```

```

FT      /*tag= c
FT      /label= bar
FT      /note= "region coding for phosphinothricin
FT      acetyltransferase"
FT      complement (883..2608)
FT      /*tag= d
FT      /label= Psu
FT      /note= "promoter of Arabidopsis Rubisco small
FT      subunit gene"
FT      complement (2659..3031)
FT      /*tag= e
FT      /label= 3'nos
FT      /note= "region containing polyA signal of nopaline
FT      synthase gene of Agrobacterium T-DNA"
FT      complement (3032..3367)
FT      /*tag= f
FT      /label= Barnase
FT      /note= "Bacillus amyloliquefaciens barnase coding
FT      region"
FT      complement (3368..4877)
FT      /*tag= g
FT      /label= PTA29
FT      /note= "promoter of stamen-specific TA29 gene of
FT      Nicotiana tabacum"
FT      4924..5216
FT      /*tag= h
FT      /label= Pnos
FT      /note= "promoter of nopaline synthase gene of
FT      Agrobacterium T-DNA"
FT      5217..5489
FT      /*tag= i
FT      /label= Barstar
FT      /note= "region coding for barstar of Bacillus
FT      amyloliquefaciens"
FT      5490..5765
FT      /*tag= j
FT      /label= 3'g7
FT      /note= "region containing polyA signal of gene 7
FT      of Agrobacterium T-DNA"
FT      complement (5840..5864)
FT      /*tag= k
FT      /label= LB
FT      /note= "left border of Agrobacterium T-DNA"
XX
XX      WO9626283-A1.
XX
XX      29-AUG-1996.
XX
XX      21-FEB-1996; 96WO-EP00722.
XX
XX      21-FEB-1995; 95EP-0400364.
XX
XX      (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX      Botterman J, Cornelissen M, Michiels F;
XX      WPI; 1996-402373/40.
XX
XX      Prodn. of male sterile plants by transforming with a chimaeric
XX      construct - comprising a male sterility DNA e.g. barnase and a
XX      co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX      for generating hybrid cultivars
XX
XX      Example 3; Page 33-3743-47; 56pp; English.
XX
XX      Plasmid pTCO113 (AAT39339) is a T-DNA vector containing a bar gene
XX      under control of the PSSU promoter, a barnase gene under control
XX      of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
XX      gene under control of the Pnos promoter. 87% Of oilseed rape
XX      plants regenerated after Agrobacterium-mediated transformation
XX      using pTCO113 were male sterile. Barnase expression disturbed the
XX      function of stamen cells leading to male sterility. Constitutive
XX      expression of barstar counteracted any low level expression of

```

```
CC barnase in non-stamen tissue.
XX
SQ Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;

Query Match      100.0%; Score 25; DB 17; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaagctagc 25
   |||||
Db 317 GGATCCCCCGATGACTAAGCTAGC 293

RESULT 14
AAD06990
ID AAD06990 standard; DNA; 5865 BP.
XX
AC AAD06990;
XX
DT 06-AUG-2001 (first entry)
XX
DE Chimeric T-DNA of plasmid pTC0113.
XX
KW T-DNA; plasmid pTC0113; transgenic Brassica plant; transformation event;
KW male-sterility gene; chimeric; tobacco; ds.
XX
OS Chimeric - Streptomyces hygroscopicus.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Bacillus amyloliquefaciens.
OS Chimeric - Nicotiana tabacum.
OS Chimeric - Agrobacterium tumefaciens.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "Right border repeat from the TL-DNA from
FT pTiB6S3"
FT misc_feature 26..53
FT /tag= b
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 54..90
FT /tag= c
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 3'UTR 98..309
FT /tag= d
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT misc_feature 310..331
FT /tag= e
FT /note= "Synthetic polylinker derived sequence"
FT CDS 332..883
FT /tag= f
FT /product= "Protein encoded by bialaphos resistance
FT gene (bar) of Streptomyces hygroscopicus"
FT promoter 884..2609
FT /tag= g
FT /note= "Promoter from the atSLA ribulose-1,5-bisphosphate
FT carboxylase small subunit gene from Arabidopsis thaliana"
FT misc_feature 2610..2659
FT /tag= h
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 2660..2920
FT /tag= i
FT /note= "Tag1 fragment from the 3' untranslated end of the
FT nopaline synthase gene (3'nos) from the T-DNA of pTiT37
FT and containing plant polyadenylation signals"
FT misc_feature 2921..2936
FT /tag= j
FT /note= "Synthetic polylinker derived sequence"
FT 3'UTR 2937..3032
FT /note= "Synthetic polylinker derived sequence"

FT /tag= k
FT /note= "The 3' untranslated region downstream from the
FT barnase coding sequence of Bacillus amyloliquefaciens"
FT CDS 3033..3368
FT /tag= l
FT /product= "Protein encoded by barnase gene from
FT Bacillus amyloliquefaciens"
FT promoter 3369..4878
FT /tag= m
FT /note= "Promoter region of the anther-specific gene
FT TA29 from Nicotiana tabacum"
FT misc_feature 4879..4924
FT /tag= n
FT /note= "Synthetic polylinker derived sequence"
FT promoter 4925..5215
FT /tag= o
FT /note= "Promoter of the nopaline synthase gene from the
FT T-DNA of pTiT37 of Agrobacterium tumefaciens"
FT misc_feature 5216..5217
FT /tag= p
FT /note= "Synthetic polylinker derived sequence"
FT CDS 5218..5490
FT /tag= q
FT /product= "Protein encoded by barstar gene of
FT Bacillus amyloliquefaciens"
FT misc_feature 5491..5530
FT /tag= r
FT /note= "Sequence from the 3' untranslated end of the
FT barstar gene from Bacillus amyloliquefaciens"
FT misc_feature 5531..5554
FT /tag= s
FT /note= "Synthetic polylinker derived sequence"
FT 3'UTR 5555..5766
FT /tag= t
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT misc_feature 5767..5773
FT /tag= u
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 5774..5810
FT /tag= v
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT misc_feature 5811..5840
FT /tag= w
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 5841..5865
FT /tag= x
FT /note= "Left border repeat from the TL-DNA from
FT pTiB6S3"

WO200131042-A2.
03-MAY-2001.
26-OCT-2000; 2000WO-EP10680.
29-OCT-1999; 99US-0430497.
(AVET ) AVENTIS CROPS SCIENCE NV.
Weston B, De Beuckeleer M;
WPI; 2001-300517/31.
Transgenic Brassica plants, seeds, cells or tissues, characterized by
harboring specific transformation events, particularly by presence of
male-sterility gene, at specific location in its genome.
Claim 1; Page 47-49; 53pp; English.
The present invention relates to a transgenic Brassica plant or its
seed, cells or tissues, characterised by harbouring a specific
```

```

CC transformation event, particularly by the presence of a male-sterility
CC gene, at a specific location in the Brassica genome. Transgenic
CC Brassica plant is useful for producing a hybrid seed by crossing the
CC transgenic plant with a male-fertile Brassica plant and harvesting the
CC hybrid seed from the transgenic Brassica plant.
CC The present sequence is chimeric r-DNA of plasmid pTCO113. This sequence
CC comprises right border repeat, left border repeat and 3' untranslated
CC region (UTR) from TL-DNA of pTiB6S3, synthetic polylinker sequences,
CC coding regions of Bialaphos resistance gene (bar) from
CC Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens
CC and barstar gene from Bacillus amyloliquefaciens and promoters of atsIA
CC ribulose-1,5-bisphosphate carboxylase small subunit gene from
CC Arabidopsis thaliana, the anther-specific gene TA29 from
CC Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTiT37
CC of Agrobacterium tumefaciens.
XX
SQ Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;

Query Match      100.0%; Score 25; DB 22; Length 5865;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccagatgactaagctagc 25
   |||||
Db 5547 ggatcccccagatgactaagctagc 5571

RESULT 15
AAD06990/c
ID AAD06990 standard; DNA; 5865 BP.
XX
AC AAD06990;
XX
DT 06-AUG-2001 (first entry)
XX
DE Chimeric T-DNA of plasmid pTCO113.
KW T-DNA; plasmid pTCO113; transgenic Brassica plant; transformation event;
KW male-sterility gene; chimeric; tobacco; ds.
XX
OS Chimeric - Streptomyces hygroscopicus.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Bacillus amyloliquefaciens.
OS Chimeric - Nicotiana tabacum.
OS Chimeric - Agrobacterium tumefaciens.
XX Chimeric - Unidentified.
FH
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "Right border repeat from the TL-DNA from
FT pTiB6S3"
FT misc_feature 26..53
FT /tag= b
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 54..90
FT /tag= c
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 3'UTR 98..309
FT /tag= d
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT misc_feature 310..331
FT /tag= e
FT /note= "Synthetic polylinker derived sequence"
FT CDS 332..883
FT /tag= f
FT /product= "Protein encoded by Bialaphos resistance
FT gene (bar) of Streptomyces hygroscopicus"
FT promoter 884..2609
FT /tag= g

```

```

FT /note= "Promoter from the atsIA ribulose-1,5-bisphosphate
FT carboxylase small subunit gene from Arabidopsis thaliana"
FT /tag= h
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 2660..2920
FT /tag= i
FT /note= "TaqI fragment from the 3' untranslated end of the
FT nopaline synthase gene (3'nos) from the T-DNA of pTiT37
FT and containing plant polyadenylation signals"
FT misc_feature 2921..2936
FT /tag= j
FT /note= "Synthetic polylinker derived sequence"
FT 3'UTR 2937..3032
FT /tag= k
FT /note= "The 3' untranslated region downstream from the
FT barnase coding sequence of Bacillus amyloliquefaciens"
FT CDS 3033..3368
FT /tag= l
FT /product= "Protein encoded by barnase gene from
FT Bacillus amyloliquefaciens"
FT promoter 3369..4878
FT /tag= m
FT /note= "Promoter region of the anther-specific gene
FT TA29 from Nicotiana tabacum"
FT misc_feature 4879..4924
FT /tag= n
FT /note= "Synthetic polylinker derived sequence"
FT promoter 4925..5215
FT /tag= o
FT /note= "Promoter of the nopaline synthase gene from the
FT T-DNA of pTiT37 of Agrobacterium tumefaciens"
FT misc_feature 5216..5217
FT /tag= p
FT /note= "Synthetic polylinker derived sequence"
FT CDS 5218..5490
FT /tag= q
FT /product= "Protein encoded by barstar gene of
FT Bacillus amyloliquefaciens"
FT misc_feature 5491..5530
FT /tag= r
FT /note= "Sequence from the 3' untranslated end of the
FT barstar gene from Bacillus amyloliquefaciens"
FT misc_feature 5531..5554
FT /tag= s
FT /note= "Synthetic polylinker derived sequence"
FT 3'UTR 5555..5766
FT /tag= t
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT misc_feature 5767..5773
FT /tag= u
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 5774..5810
FT /tag= v
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT misc_feature 5811..5840
FT /tag= w
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 5841..5865
FT /tag= x
FT /note= "Left border repeat from the TL-DNA from
FT pTiB6S3"
XX
WO200131042-A2.
PD 03-MAY-2001.
XX
PF 26-OCT-2000; 2000WO-EP10680.
XX
PR 29-OCT-1999; 99US-0430497.
XX

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:13 ; Search time 301.6 Seconds  
(without alignments)  
18.773 Million cell updates/sec

Title: US-09-698-903B-5

Perfect score: 25

Sequence: 1 ggatccccgatgagctagctagc 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	25	100.0	1303	3	US-08-894-440-2
c 2	25	100.0	4946	3	US-08-817-188-1
c 3	25	100.0	5560	3	US-08-817-188-5
4	25	100.0	5864	3	US-08-894-440-4
c 5	25	100.0	5864	3	US-08-894-440-4
6	25	100.0	6548	3	US-08-894-440-1
7	25	100.0	6548	3	US-08-817-188-2
8	24	96.0	3200	1	US-08-453-104-23
9	24	96.0	3200	2	US-08-694-824-23
10	24	96.0	3201	1	US-08-453-104-22
11	24	96.0	3201	2	US-08-694-824-22
12	24	96.0	7566	2	US-08-232-016-23
13	24	96.0	7639	2	US-08-232-016-22
c 14	24	96.0	7811	2	US-08-549-580A-5
c 15	17	68.0	29	2	US-08-232-016-11
16	17	68.0	3153	4	US-09-080-625-3
17	17	68.0	3336	4	US-09-080-625-2
18	17	68.0	3694	4	US-09-080-625-5
19	17	68.0	3877	4	US-09-080-625-4
c 20	17	68.0	24595	6	5428147-1
c 21	16.6	66.4	6201	2	US-08-790-912-1
22	16	64.0	54	2	US-08-503-226B-32
23	16	64.0	54	3	US-08-721-458B-32
c 24	16	64.0	3065	2	US-08-209-521-28
25	15.4	61.6	1186	1	US-08-084-121-2
26	15.4	61.6	1186	1	US-08-478-015-2
27	15.4	61.6	1186	3	US-08-475-975-2

28	15.4	61.6	1186	3	US-09-084-889-2	Sequence 2, Appli
29	15.4	61.6	1859	3	US-08-691-563C-46	Sequence 46, Appl
c 30	15.4	61.6	2143	2	US-08-656-177A-1	Sequence 1, Appli
c 31	15.4	61.6	2143	3	US-09-256-797-1	Sequence 1, Appli
c 32	15.4	61.6	10281	2	US-08-816-155B-1	Sequence 1, Appli
c 33	15.4	61.6	10281	3	US-09-079-587-1	Sequence 1, Appli
c 34	15.2	60.8	43795	3	US-08-742-185-101	Sequence 101, App
35	15	60.0	1642	1	US-07-996-772A-1	Sequence 1, Appli
36	15	60.0	1642	1	US-08-446-822-1	Sequence 1, Appli
37	15	60.0	1642	5	PCT-US93-12586-1	Sequence 1, Appli
c 38	15	60.0	2352	2	US-08-922-837-1	Sequence 1, Appli
c 39	15	60.0	2352	4	US-09-351-550-1	Sequence 1, Appli
c 40	15	60.0	7431	4	US-09-306-998-2	Sequence 2, Appli
c 41	14.8	59.2	1377	1	US-08-112-817C-1	Sequence 1, Appli
42	14.8	59.2	1710	4	US-09-068-960-14	Sequence 14, Appli
43	14.6	58.4	376	3	US-08-985-950-3	Sequence 3, Appli
c 44	14.6	58.4	699	4	US-08-998-416-591	Sequence 591, App
c 45	14.4	57.6	60	4	US-09-171-945-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1  
US-08-894-440-2  
; Sequence 2, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894,440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1303  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-ECORI  
; OTHER INFORMATION: fragment of pT588  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(35)  
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36)..(694)  
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain  
; OTHER INFORMATION: CM1841 (P35S)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (695)..(967)  
; OTHER INFORMATION: region coding for barstar of Bacillus  
; OTHER INFORMATION: amylioliquefaciens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (968)..(1287)  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1288)..(1303)  
; OTHER INFORMATION: polylinker of pGEM2  
; US-08-894-440-2

Query Match 100.0%; Score 25; DB 3; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ggatccccgatgagctagctagc 25

Db 1024 ggatcccccgatgagctaagctagc 1048  
 |||

RESULT 2  
 US-08-817-188-1/c  
 ; Sequence 1, Application US/08817188  
 ; Patent No. 6074876  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DE BLOCK, MARC  
 ; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 ; FILE REFERENCE: 2121-0127P  
 ; CURRENT APPLICATION NUMBER: US/08/817,188  
 ; CURRENT FILING DATE: 1997-05-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
 ; EARLIER FILING DATE: 1996-07-31  
 ; EARLIER APPLICATION NUMBER: EP 95401844.6  
 ; EARLIER FILING DATE: 1995-08-04  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 4946  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
 ; OTHER INFORMATION: plasmid pTHW107  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement(1)..(25)  
 ; OTHER INFORMATION: T-DNA right border (RB)  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((97)..(330))  
 ; OTHER INFORMATION: 3' g7: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
 ; OTHER INFORMATION: T-DNA  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((331)..(882))  
 ; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
 ; OTHER INFORMATION: transferase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((883)..(2608))  
 ; OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
 ; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((2658)..(3031))  
 ; OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
 ; OTHER INFORMATION: gene of Agrobacterium T-DNA  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((3032)..(3367))  
 ; OTHER INFORMATION: barnase: region coding for barnase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((3368)..(4876))  
 ; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana  
 ; OTHER INFORMATION: tabacum  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((4922)..(4946))  
 ; OTHER INFORMATION: LB: T-DNA left border  
 US-08-817-188-1

Query Match 100.08; Score 25; DB 3; Length 4946;  
 Best Local Similarity 100.08; Pred. No. 0.0014;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaagctagc 25  
 |||  
 Db 317 GGATCCCCGATGAGCTAAGCTAGC 293

RESULT 3  
 US-08-817-188-5/c  
 ; Sequence 5, Application US/08817188  
 ; Patent No. 6074876  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DE BLOCK, MARC  
 ; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 ; FILE REFERENCE: 2121-0127P  
 ; CURRENT APPLICATION NUMBER: US/08/817,188  
 ; CURRENT FILING DATE: 1997-05-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
 ; EARLIER FILING DATE: 1996-07-31  
 ; EARLIER APPLICATION NUMBER: EP 95401844.6  
 ; EARLIER FILING DATE: 1995-08-04  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 5  
 ; LENGTH: 5560  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
 ; OTHER INFORMATION: plasmid pTHW142  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(25)  
 ; OTHER INFORMATION: PTIB6S3  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (84)..(296)  
 ; OTHER INFORMATION: 3' g7: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
 ; OTHER INFORMATION: T-DNA  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (318)..(869)  
 ; OTHER INFORMATION: bar: region coding for phosphinotricin  
 ; OTHER INFORMATION: acetyltransferase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (830)..(2760)  
 ; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit  
 ; OTHER INFORMATION: gene of Arabidopsis thaliana  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (2765)..(3058)  
 ; OTHER INFORMATION: 3' untranslated region of the CaMV 35S transcript  
 ; OTHER INFORMATION: containing polyadenylation signals  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (3059)..(5056)  
 ; OTHER INFORMATION: uidA: region coding for beta-glucuronidase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4483)..(4671)  
 ; OTHER INFORMATION: IV2: region corresponding to the second intron of  
 ; OTHER INFORMATION: the ST-LS1 gene  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (5087)..(5502)  
 ; OTHER INFORMATION: P35S: 35S promoter region of CaMV  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (5533)..(5560)  
 ; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from  
 ; OTHER INFORMATION: PTIB6S3  
 ; FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; US-08-817-188-5

Query Match      100.0%; Score 25; DB 3; Length 5560;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatccccgatgagctaaagctagc 25
    |||||
Db 304 GGATCCCCGATGAGCTAAGCTAGC 280

RESULT 4
US-08-894-440-4
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; TYPE: DNA
; LENGTH: 5864
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTCO113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))

```

```

; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Phos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
; US-08-894-440-4

Query Match      100.0%; Score 25; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatccccgatgagctaaagctagc 25
    |||||
Db 5546 ggatccccgatgagctaaagctagc 5570

RESULT 5
US-08-894-440-4/c
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; TYPE: DNA
; LENGTH: 5864
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTCO113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4

```

```

Query Match      100.0%; Score 25; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
    ||||||||||||||||||||||||
Db 317 GGATCCCCGATGAGCTAGCTAGC 293

```

```

RESULT 6
US-08-894-440-1
; Sequence 1, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTS174
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2003)
; OTHER INFORMATION: pUC19 derived vector sequences (vector)
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2283))
; OTHER INFORMATION: 3' nos: region containing polyadenylation signal
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2284)..(2624))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: promoter of the stamen-specific B1 gene of rice
; OTHER INFORMATION: (PEI)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5730)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus (P35S)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5711)..(6262)
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: region containing polyadenylation signal fo gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
US-08-894-440-1

```

```

Query Match      100.0%; Score 25; DB 3; Length 6548;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
    ||||||||||||||||||||||||
Db 6277 ggatcccccgatgagctaaagctagc 6301

```

```

RESULT 7
US-08-817-188-2
; Sequence 2, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTS172
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2288))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2289)..(2624))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:

```

```
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: PE1: promoter region of E1 gene of rice
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5170)
; OTHER INFORMATION: P35S: 35S promoter region of Cauliflower Mosaic
; OTHER INFORMATION: Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5711)..(6262))
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
US-08-817-188-2

Query Match          100.0%; Score 25; DB 3; Length 6548;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaaagctagc 25
|||||
Db 6277 ggatcccccgatgactaaagctagc 6301

RESULT 8
US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,104
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
US-08-453-104-23

Query Match          96.0%; Score 24; DB 1; Length 3200;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gatcccccgatgactaaagctagc 25
|||||
Db 2879 GATCCCCGATGAGCTAAGCTAGC 2902

RESULT 9
US-08-694-824-23
; Sequence 23, Application US/08694824
; Patent No. 5877306
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,824
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
```

; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; - OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 96.0%; Score 24; DB 2; Length 3200;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatccccgatgactaagctagc 25  
|||||  
Db 2879 GATCCCCGATGAGCTAAGCTAGC 2902

RESULT 10  
US-08-453-104-22  
; Sequence 22, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; - OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 96.0%; Score 24; DB 1; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 0.004;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 gatccccgatgactaagctagc 25  
|||||  
Db 2952 GATCCCCGATGAGCTAAGCTAGC 2975

RESULT 11  
US-08-694-824-22  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; - OTHER INFORMATION: wherein N is not known."  
US-08-694-824-22

Query Match 96.0%; Score 24; DB 2; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatccccgatgactaagctagc 25  
|||||  
Db 2952 GATCCCCGATGAGCTAAGCTAGC 2975

RESULT 12



US-08-232-016-23  
; Sequence 23, Application US/08232016  
; Patent No. 5952547  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; APPLICANT: VAN AARSEN, Roel  
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN  
; TITLE OF INVENTION: PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.016  
; FILING DATE: 03-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 91402920.2  
; FILING DATE: 30-OCT-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 92400820.4  
; FILING DATE: 25-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K  
; REGISTRATION NUMBER: P39,300  
; REFERENCE/DOCKET NUMBER: 010830-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7566 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: plasmid DNA designated as pPS0212  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1785  
; OTHER INFORMATION: /note= "Coding region of a  
; OTHER INFORMATION: truncated modified bt2 (cryIab) gene, also designated as the  
; OTHER INFORMATION: cryIab6 gene."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1793..2026  
; OTHER INFORMATION: /note= "3' regulatory sequence  
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu  
; OTHER INFORMATION: T-DNA gene 7."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2396..2921  
; OTHER INFORMATION: /note= "35S promoter sequence  
; OTHER INFORMATION: derived from Cauliflower mosaic virus."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2922..3581  
; OTHER INFORMATION: /note= "Coding sequence of  
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."  
; FEATURE:

; NAME/KEY: misc\_feature  
; LOCATION: 3582..4407  
; OTHER INFORMATION: /note= "3' regulatory sequence  
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacte  
; OTHER INFORMATION: T-DNA octopine synthase gene."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 5600..6457  
; OTHER INFORMATION: /note= "sequence complementary to  
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 7071..7566  
; OTHER INFORMATION: /note= "TR1' and TR2 promoter  
; OTHER INFORMATION: derived from Agrobacterium T-DNA (with modified leader wit  
; OTHER INFORMATION: respect to sequence of pUD884 of SEQ ID NO. 22."  
US-08-232-016-23  
  
Query Match 96.0%; Score 24; DB 2; Length 7566;  
Best Local Similarity 100.0%; Pred. No. 0.0045;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 gatccccgatgagctaagctagc 25  
|||||  
Db 1808 GATCCCCGATCAGCTAAGCTAGC 1831  
  
RESULT 13  
US-08-232-016-22  
; Sequence 22, Application US/08232016  
; Patent No. 5952547  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; APPLICANT: VAN AARSEN, Roel  
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN  
; TITLE OF INVENTION: PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.016  
; FILING DATE: 03-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 91402920.2  
; FILING DATE: 30-OCT-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 92400820.4  
; FILING DATE: 25-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K  
; REGISTRATION NUMBER: P39,300  
; REFERENCE/DOCKET NUMBER: 010830-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7566 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: plasmid DNA designated as pPS0212  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1785  
; OTHER INFORMATION: /note= "Coding region of a  
; OTHER INFORMATION: truncated modified bt2 (cryIab) gene, also designated as the  
; OTHER INFORMATION: cryIab6 gene."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1793..2026  
; OTHER INFORMATION: /note= "3' regulatory sequence  
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu  
; OTHER INFORMATION: T-DNA gene 7."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2396..2921  
; OTHER INFORMATION: /note= "35S promoter sequence  
; OTHER INFORMATION: derived from Cauliflower mosaic virus."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2922..3581  
; OTHER INFORMATION: /note= "Coding sequence of  
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."  
; FEATURE:

```
; LENGTH: 7639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE: plasmid DNA designated as pJD884
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1869
;   OTHER INFORMATION: /note= "coding region of a
;   OTHER INFORMATION: truncated bc2 (cryIab) gene, also designated as the bc884 gen
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1877..2110
;   OTHER INFORMATION: /note= "3' regulatory sequence
;   OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
;   OTHER INFORMATION: T-DNA gene 7."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2480..3005
;   OTHER INFORMATION: /note= "35S promoter sequence
;   OTHER INFORMATION: derived from Cauliflower mosaic virus."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3006..3665
;   OTHER INFORMATION: /note= "coding sequence of
;   OTHER INFORMATION: chloramphenicol acetyl transferase gene."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3666..4491
;   OTHER INFORMATION: /note= "3' regulatory sequence
;   OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
;   OTHER INFORMATION: T-DNA octopine synthase gene."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 5684..6541
;   OTHER INFORMATION: /note= "Sequence complementary to
;   OTHER INFORMATION: Patent No. 5952547
;   OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 7155..7639
;   OTHER INFORMATION: /note= "Tr1' and TR2' promoter
;   OTHER INFORMATION: derived from Agrobacterium T-DNA."
; US-08-232-016-22
```

```
Query Match          96.0%; Score 24; DB 2; Length 7639;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 gatccccgatgagctaagctagc 25
Db 1892 GATCCCCGATGAGCTAAGCTAGC 1915
```

```
RESULT 14
US-08-549-680A-5/c
; Sequence 5, Application US/08549680A
; Patent No. 5962768
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, MARCUS
; APPLICANT: REYNAERTS, ARLETTE
; APPLICANT: GOSSELE, VERNIQUE
; APPLICANT: VAN AARSEN, ROEL
; TITLE OF INVENTION: MARKER GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
```

```
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,680A
; FILING DATE: 16 JANUARY 1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-0111P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7811 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (synthetic)
; FEATURE:
;   NAME/KEY: misc_recomb
;   LOCATION: 1..7811
;   OTHER INFORMATION: /label= vector pTRVA3
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 194..218
;   OTHER INFORMATION: /note= "T-DNA right border"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 484..684
;   OTHER INFORMATION: /note= "the 3' end formation and
;   OTHER INFORMATION: polyadenylation region of T-DNA gene 7"
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 729..1340
;   OTHER INFORMATION: /note= "the aac(6') coding
;   OTHER INFORMATION: sequence"
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: 1341..1756
;   OTHER INFORMATION: /label= 35S promoter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3001..3023
;   OTHER INFORMATION: /note= "T-DNA left border
;   OTHER INFORMATION: sequences"
; US-08-549-680A-5
```

```
Query Match          96.0%; Score 24; DB 2; Length 7811;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 gatccccgatgagctaagctagc 25
Db 702 GATCCCCGATGAGCTAAGCTAGC 679
```

```
RESULT 15
US-08-232-016-11/c
; Sequence 11, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
```

APPLICANT: DOCKX, Jan  
APPLICANT: VAN AARSEN, Roel  
TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN  
TITLE OF INVENTION: PLANT CELLS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232.016  
FILING DATE: 03-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91402920.2  
FILING DATE: 30-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 92400820.4  
FILING DATE: 25-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K  
REGISTRATION NUMBER: P39,300  
REFERENCE/DOCKET NUMBER: 010830-049  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: oligonucleotide, designated as PS59  
US-08-232-016-11

Query Match 68.0%; Score 17; DB 2; Length 29;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 gatcccccgatgagcta 18  
|||||  
Db 17 GATCCCCCGATGAGCTA 1

Search completed: February 25, 2002, 18:05:15  
Job time: 18563 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:20:57 ; Search time 8261.74 Seconds  
(without alignments)  
32.517 Million cell updates/sec

Title: US-09-698-903B-5  
Perfect score: 25  
Sequence: 1 ggatccccgatgagctaagctagc 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estfun: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estom: \*  
5: em\_estpl: \*  
6: em\_estba: \*  
7: em\_estro: \*  
8: em\_estov: \*  
9: em\_htc: \*  
10: gb\_est1: \*  
11: gb\_est2: \*  
12: gb\_htc: \*  
13: gb\_gss: \*  
14: em\_gss\_fun: \*  
15: em\_gss\_hum: \*  
16: em\_gss\_inv: \*  
17: em\_gss\_pln: \*  
18: em\_gss\_pro: \*  
19: em\_gss\_rod: \*  
20: em\_gss\_vrt: \*  
21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.2	72.8	498	13 AQ791259	HS_5498_A
c 2	18.2	72.8	562	13 AZ028032	RPCI-23-3
c 3	17.6	70.4	303	10 AW862033	RC3-CT034
c 4	17.6	70.4	312	10 BB453584	BB453584
c 5	17.6	70.4	454	11 N24161	N24161.yx96d12.s1
c 6	17.6	70.4	497	11 BF496895	BF496895.AT10901.5
c 7	17.6	70.4	510	11 BI038578	BI038578.MR4-NF014
c 8	17.6	70.4	587	11 BF877912	BF877912.MR0-ET010
c 9	17.6	70.4	605	10 BE639934	BE639934.946044B07
c 10	17.6	70.4	611	13 AQ834638	AQ834638.HS_5442_A
c 11	17.6	70.4	924	13 CNS038XY	AL233215 Tetraodon
c 12	17.2	68.8	295	10 BE120999	UI-R-CA0-

13	17.2	68.8	326	10	BB499943	BB499943
c 14	17.2	68.8	336	11	BF410724	UI-R-CA0-
c 15	17.2	68.8	368	13	AQ867402	nbeb0031F
c 16	17.2	68.8	387	10	AW501051	UI-HF-BP0
c 17	17.2	68.8	398	10	AI183978	qdg9g01.x
c 18	17.2	68.8	407	10	AA429095	zva49d03.f
c 19	17.2	68.8	410	10	AW581963	MR4-ST012
c 20	17.2	68.8	462	11	BI134877	UI-M-BH03-
c 21	17.2	68.8	501	11	BG073283	H3120G12-
c 22	17.2	68.8	549	13	TA311E12Q	T. brucei
c 23	17.2	68.8	554	11	BI133671	UI-M-BH3-
c 24	17.2	68.8	554	13	AQ509694	nbxd0096F
c 25	17.2	68.8	568	10	BE048724	hr52h07.x
c 26	17.2	68.8	574	10	BE026535	db27g06.x
c 27	17.2	68.8	583	11	BI134114	UI-M-BH3-
c 28	17.2	68.8	585	10	BE326762	hr64e06.x
c 29	17.2	68.8	622	10	AW548994	L0045H03-
c 30	17.2	68.8	627	10	BE535820	601060510
c 31	17.2	68.8	630	11	BG470582	602511491
c 32	17.2	68.8	707	11	BG334339	602461858
c 33	17.2	68.8	729	10	BE394825	601312066
c 34	17.2	68.8	755	10	BE543444	601070751
c 35	17.2	68.8	768	11	BG678254	602624567
c 36	17.2	68.8	781	10	AI344042	tc01905.x
c 37	17.2	68.8	812	11	BG035724	602325804
c 38	17.2	68.8	825	11	BG166218	602340976
c 39	17.2	68.8	894	11	BF301141	602029166
c 40	17.2	68.8	916	11	BF797773	602257530
c 41	17.2	68.8	992	13	CNS0333U	AL225651 Tetraodon
c 42	17.2	68.8	1498	11	BG476758	602524680
c 43	17.2	68.8	1649	12	AK016762	Mus muscu
c 44	17	68.0	251	10	BE419550	WWS014.E1
c 45	17	68.0	400	10	BE119925	UI-R-CA0-

#### ALIGNMENTS

RESULT 1  
LOCUS AQ791259 498 bp DNA GSS 03-AUG-1999  
DEFINITION HS\_5498\_A2\_B10\_77A RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=1074 Col=20 Row=C, DNA sequence.  
ACCESSION AQ791259  
VERSION AQ791259.1 GI:5698806  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 498)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pletier de Jong  
(pletier@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.hscsc.washington.edu  
Plate: 1074 row: C column: 20  
Seq primer: T7

```

Class: BAC ends
High quality sequence stop: 498.
Location/Qualifiers
  1. 498
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="Plate-1074 Col=20 Row=C"
  /clone_lib="RPCI-11 Human Male BAC Library"
  /sex="male"
  /note="vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT      166 a   90 c   85 g   151 t       6 others
ORIGIN

Query Match      72.8%; Score 18.2; DB 13; Length 498;
Best Local Similarity 87.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atccccgatgagctagctagc 25
||||| ||||| ||||| |||||
Db 414 ATCCCCGATGCTTAAGATAGC 436

RESULT 2
AZ028032/c
LOCUS      562 bp      DNA      GSS      25-FEB-2000
DEFINITION  RPCI-23-373M15-TV RPCI-23 Mus musculus genomic clone RPCI-23-373M15
, DNA sequence.
ACCESSION  AZ028032
VERSION    AZ028032.1 GI:7103416
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 562)
Zhaoh,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., Megann,S., Isegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-373M15.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 373 row: M column: 15
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
  1. .562
  /organism="Mus musculus"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone_lib="RPCI-23-373M15"
  /clone_lib="RPCI-23"
  /sex="Female"
  /lab_host="DH10B"
  /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies).
BASE COUNT      122 a   122 c   163 g   153 t       2 others
ORIGIN

Query Match      72.8%; Score 18.2; DB 13; Length 562;
Best Local Similarity 87.0%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atccccgatgagctagctagc 25
||||| ||||| ||||| |||||
Db 247 ATCCACCAATGAGCTTAGCTAGC 225

RESULT 3
AW862033/c
LOCUS      303 bp      mRNA      EST      19-MAY-2000
DEFINITION  RC3-CT0347-110300-014-a10 CT0347 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW862033
VERSION    AW862033.1 GI:7957726
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 303)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=42-RC3-CT0347-110
300-014-a10&t3=2000-03-11&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 303.
Location/Qualifiers
  1. 303
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="CT0347"
  /dev_stage="Adult"
  /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      104 a   60 c   62 g   77 t
ORIGIN

Query Match      70.4%; Score 17.6; DB 10; Length 303;

```

Best Local Similarity 83.3%; Pred. No. 5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ggtccccgtagctaaagctagc 25  
|||||  
Db 29 GAGTCCTTATGAGCTAAGCTAGC 6

## RESULT 4

BB453584 312 bp mRNA EST 21-JUL-2000  
BB453584 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D130023P09 3' similar to U28724 Mus  
musculus DNA mismatch repair (PMS2) mRNA, mRNA sequence.

ACCESSION BB453584.1 GI:9349077

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 312)

## REFERENCE

## AUTHORS

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci  
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,  
Y., Shigenoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suganara, Y.,  
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya  
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,  
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino  
M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)

## TITLE

## JOURNAL

## COMMENT

Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki  
Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## FEATURES

## source

Location/Qualifiers  
1..312  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130023P09"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGCGGCGCACTCGAGTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',  
GAGAGAGATTCTCGACTTAATAATTAATCCCGCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."  
BASE COUNT 81 a 74 c 60 g 96 t 1 others  
ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 312;  
Best Local Similarity 83.3%; Pred. No. 5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggatccccgtagctaaagctagc 24  
|||||  
Db 239 GGATCTCCCAAGTAGCTAAGCTAG 216

## RESULT 5

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## 1..454

## /organism="Homo sapiens"

## /db\_xref="GDB:3879233"

## /db\_xref="taxon:9606"

## /clone="IMAGE:269591"

## /clone\_lib="Soares melanocyte 2NbHM"

## /sex="Male"

## /tissue\_type="melanocyte"

## /lab\_host="DH10B (ampicillin resistant)"

/note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAAGTGGCGCGCGAGTTTCTTTTCTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector

(Pharmacía). Library constructed by Bento Soares and M.Fátima Bonafide. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino.  
 BASE COUNT 173 a 81 c 61 g 139 t  
 ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 454;  
 Best Local Similarity 83.3%; Pred. No. 5.3e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 ggatccccgatgactaagctag 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 448 GGATCCCCGAGGTGAGTAGCTAG 425

RESULT 6  
 LOCUS BF496895 497 bp mRNA EST 19-APR-2001  
 DEFINITION AT10901.5prime AT Drosophila melanogaster adult testes pOTB7  
 Drosophila melanogaster cDNA clone AT10901 5 similar to CG7235;  
 FBan0007235 'heat shock protein' located on: 2L 25F4-25F4;  
 04/08/2001, mRNA sequence.  
 ACCESSION BF496895  
 VERSION BF496895.2 GI:13694366  
 KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 497)

REFERENCE  
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Parfan, D., Frise, E., George, R., Gonzalez, M., Guarín, H., Harris, N., Li, P., Liao, G., Miranda, A., Mista, S., Muncall, C.J., Nunoo, J., Pacleb, J., Parades, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celisner, S., and Rubin, G.M.  
 TITLE BGP/HMI AT Drosophila EST Project  
 JOURNAL Unpublished (2000)  
 COMMENT On Dec 6, 2000 this sequence version replaced gi:11580196.  
 Contact: Stapleton, M.  
 BDGP

Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
 hit genomic AE003610; arm:2L [5357975,5614606]  
 estimated-cyto:25E2-25F4; 04/08/2001  
 Plate: AT.109 row: A column: 1  
 High quality sequence stop: 455.  
 Location/Qualifiers  
 1..497

FEATURES  
 source  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="AT10901"  
 /dev\_stage="Adult"  
 /note="Organ: testes; Vector: pOTB7; Site: 1: EcoRI; Site: 2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."  
 125 a 123 c 159 g 90 t

Query Match 70.4%; Score 17.6; DB 11; Length 497;  
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 gatccccgatgactaagctagc 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 452 GATCGCGGAGGAGTTATGCTAGC 475

RESULT 7  
 LOCUS BI038578 510 bp mRNA EST 14-JUN-2001  
 DEFINITION MR4-NT0143-220101-004-h09 NT0143 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BI038578  
 VERSION BI038578.1 GI:14445204  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 510)  
 REFERENCE  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&t2=MR4-NT0143-220101-004-h09&t3=2001-01-22&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 196  
 High quality sequence stop: 396.  
 Location/Qualifiers  
 1..510  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NT0143"  
 /dev\_stage="Adult"  
 /note="Organ: nervous tumor; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 144 a 119 c 110 g 137 t

BASE COUNT 144 a 119 c 110 g 137 t  
 ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 510;  
 Best Local Similarity 83.3%; Pred. No. 5.4e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggatccccgatgactaagctag 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 503 GGATCCCCGATGAGTCAATCTTG 480

RESULT 8  
 LOCUS BF877912 587 bp mRNA EST 17-JAN-2001



```

DEFINITION MR0-ET0109-171100-001-a01 ET0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF877912
VERSION BF877912.1 GI:12268042
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 587)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., Geolvelira,P., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202863
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-ET0109-
171100-001-a01&t3=2000-11-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 19
High quality sequence stop: 521.
FEATURES
source
Location/Qualifiers
1..587
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0109"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 149 a 131 c 164 g 143 t
ORIGIN
1 ggatcccccgatgaactaagctag 24
|||||
15 GGATCCCCCGATGAGTAATCTTG 38

Query Match 70.4%; Score 17.6; DB 11; Length 587;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ggatcccccgatgaactaagctag 24
|||||
15 GGATCCCCCGATGAGTAATCTTG 38

Db 15 GGATCCCCCGATGAGTAATCTTG 38

RESULT 9
LOCUS BE639934 605 bp mRNA EST 30-AUG-2000
DEFINITION 94604B07.y1 946 - tassal primordium prepared by Schmidt lab zea
mays cDNA, mRNA sequence.
ACCESSION BE639934
VERSION BE639934.1 GI:9953351
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 605)
AUTHORS Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946044 row: B column: 07.
Location/Qualifiers
1..605
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassal primordium prepared by Schmidt
lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XlOIR"
/note="Organ: tassels; Vector: HybrizAP; Site_1: EcoRI;
Site_2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybrizAP. Sample insert size range was 350 bp
to 3 Kb with a 1 Kb average."
BASE COUNT 138 a 160 c 186 g 120 t 1 others
ORIGIN
1 ggatcccccgatgaactaagctagc 25
|||||
533 GAATCCCCGNTGACTAGGCAAGC 557

Query Match 70.4%; Score 17.6; DB 10; Length 605;
Best Local Similarity 80.0%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ggatcccccgatgaactaagctagc 25
|||||
533 GAATCCCCGNTGACTAGGCAAGC 557

Db 533 GAATCCCCGNTGACTAGGCAAGC 557

RESULT 10
LOCUS A0834638 611 bp DNA GSS 27-AUG-1999
DEFINITION HS_5442_AL_H01_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1018 Col=1 Row=0, DNA sequence.
ACCESSION A0834638
VERSION A0834638.1 GI:5800700
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 611)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999).
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

```

```

FEATURES
  source
    1..924
      Location/Qualifiers
        nclup://www.genoscope.cns.fr/Tetraodon.
        /organism="Tetraodon nigroviridis"
        /db_xref="taxon:99883"
        /clones="006G08"

```

TAG\_ISSUE=HYPOCHALMUS  
TAG\_SEQ=GATGC#

```

BASE COUNT      75 a      67 c      72 g      81 t
ORIGIN

Query Match      68.8%; Score 17.2; DB 10; Length 295;
Best Local Similarity 86.4%; Pred. No. 7.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggatcccgatgagctaaagt 22
||||| ||||| ||||| |||||
Db 237 GGATCCCGACGACGAGACT 216

RESULT 13
LOCUS      BB499943      326 bp      mRNA      EST      26-JUL-2000
DEFINITION BB499943 RIKEN full-length enriched, 0 day neonate kidney Mus
musculus cDNA clone D630023P14 3', mRNA sequence.
ACCESSION  BB499943
VERSION     BB499943
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Arakawa.T., Carninci
P., Endo.T., Fukuda.S., Fukunishi.Y., Hara.A., Hayatsu.N.,
Hirozane.T., Horii.F., Ishii.Y., Ishikawa.J., Ishikawa.T., Itoh.M.,
Izawa.M., Kadota.K., Kagawa.I., Kai.C., Kawai.J., Kikuchi.N.,
Kiyosawa.H., Kojima.Y., Kondo.S., Koya.S., Kurihara.C., Kusakabe.M.,
Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H., Okazaki.Y.,
Ono.T., Owa.C., Saito.H., Sakai.C., Sato.K., Shibata.K., Shibata
Y., Shigemoto.Y., Shingagawa.A., Shiraki.T., Sogabe.Y., Sugahara.Y.,
Suzuki.H., Suzuki.H., Tagawa.A., Takahashi.F., Tomimaga.N., Toya
T., Tsunoda.Y., Watahiki.A., Watanabe.S., Yamamura.T., Yamanaka.I.,
Yano.R., Yasunishi.A., Yokota.T., Yoshida.K., Yoshiki.A., Yoshino
M., Muramatsu.M. and Hayashizaki.Y.
RIKEN Mouse ESTs (Konno.H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES             Location/Qualifiers
     source           1..326
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D630023P14"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     kidney"
                     /tissue_type="kidney"
                     /dev_stage="0 day neonate"

/lab_host="DH10B"
/note="site_1: Sail; site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCGAGTTAAATTAATTCCTCCCTCCCTCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT      95 a      68 c      60 g      103 t
ORIGIN

Query Match      68.8%; Score 17.2; DB 10; Length 326;
Best Local Similarity 86.4%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tccccgatgagctaaagtacg 25
||||| ||||| ||||| |||||
Db 107 TCCCCGATGAGCGAGCTGCG 128

RESULT 14
LOCUS      BF410724/c      336 bp      mRNA      EST      28-NOV-2000
DEFINITION UI-R-CA0-bmc-b-04-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
UI-R-CA0-bmc-b-04-0-UI 3', mRNA sequence.
ACCESSION  BF410724
VERSION     BF410724
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 336)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
discovey
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   9704477
COMMENT   Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mssoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dr track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dr track served to identify it as a clone from the
normalized thalamus library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES             Location/Qualifiers
     source           1..336
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA0-bmc-b-04-0-UI"
                     /clone_lib="UI-R-CA0"
                     /lab_host="DH10B (Life Technologies)"

```

/notes=Vector: pT773D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not 1; Site\_2: Eco RI; The UI-R-CAO  
 library is a subtracted library derived from the following  
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
 & midbrain, cerebral cortex, corpus striatum, testis, and  
 hippocampus. For a detailed description of the library  
 from which this clone was derived, please visit our web  
 site at [ratest.eng.utoro.edu](http://ratest.eng.utoro.edu). The subtraction has been  
 previously described in (Bonaldo, Lennon and Soares,  
 Genome Research 6:791-806, 1996)  
 TAG LIB=UI-R-CAO  
 TAG TISSUE=thalamus  
 TAG\_SEQ=GATCC\*

BASE COUNT	82 a	73 c	85 g	96 t
ORIGIN				
	TAG_SEQ=GATCG*			

Query Match 68.8%; Score 17.2; DB 11; Length 336;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indexes 0

Qy 1 ggatccccgatgagctaaagct 22  
|||  
Db 238 GGATCCCCGACCGGAAGCT 217

Db 238 GGATCCCCGACCAGCGAAGCT 217

RESULT 15

AQ867402	368 bp	DNA
LOCUS	nbs003lrf05f	CUGI Rice BAC Library (EcoRI) Oryza sativa genomic clone nbs003lrf05f, DNA sequence.
DEFINITION	AQ867402	GSS.
ACCESSION	AQ867402.1	GI:6217859
VERSION	GSS.	Oryza sativa.
KEYWORDS		Oryza sativa.
SOURCE		Oryza sativa.
ORGANISM		Oryza sativa.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 368)

Wing, R.A. and Dean, R.A.  
A PAC End Source of

**THE JOURNAL OF THE SAC AND SEQUENCING FRAMEWORK TO SEQUENCE THE RICE GENOME UNPUBLISHED (1998)**

COMMENT  
Contact: Wing RA

Clemson University  
Clemson University  
Clemson University

Clemson University  
100 Jordan Hall, Clemson, SC 29634 USA

Tel: 864 656 7288

Fax: 864 656 4293

Smart: TWINGEC1EM3  
Seq primer: TAAATAC

Class: BAC ends

High quality sequence sta

## FEATURES

source  
1. .368

/organism="Oryza sativa"

```
/strain="Japonica"  
/cultivar="Nipponbare"
```

```
/db_xref="taxon:4530
```

```
/clone="n
```

```
/clone_lib=-CUG1 Rice BAC Library (ECORI)"
/tissue type="Leaf"
```

```

/lab_host="E. coli

```

```
/note="Vector: pBACIndi
```

rice is the most important food crop in the world population.

populated areas of the

on rice as their primary source of carbohydrates

monocotyledonous Fice is a haploid genome equivalent

Earle, 1991). The relative

times larger than that of Arabidopsis, makes it suitable

for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation of a particular sequence with a probability of 99.9 %. Three high density filters, each containing 16,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu))."

BASE COUNT	101 a	85 c	78 g	104 t
ORIGIN				

```
Query Match      68.8%; Score 17.2; DB 13; Length 368;
Best Local Similarity 86.4%; Pred. NO. 7.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0
```

Qy 3 atccccgatgagctaagctag 24

Db  
34 ATCAACCAATGAGCTCAGCTAG 55

Search completed: February 25, 2002, 17:21:01  
Job time: 16154 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:00:10 ; Search time 2331.3 Seconds  
(without alignments)  
155.680 Million cell updates/sec

Title: US-09-698-903B-6

Perfect score: 22  
Sequence: 1 tcctctacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_ov.\*

21: em\_or.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	100.0	22	6	AX127753	AX127753 Sequence
2	22	100.0	22	6	AX172462	AX172462 Sequence
3	22	100.0	249	12	ARGMTUB	X05579 Soybean bet
4	22	100.0	831	1	ATRN7	V00090 Agrobacteri
5	22	100.0	878	1	ATTDNA	X00431 Agrobacteri
6	22	100.0	1037	6	A10942	A10942 Nucleotide
7	22	100.0	1085	6	A10939	A10939 Nucleotide
8	22	100.0	1160	6	A10943	A10943 Nucleotide
9	22	100.0	1166	6	A10941	A10941 Nucleotide
10	22	100.0	1186	6	A18051	A18051 DNA used as
11	22	100.0	1186	6	AR095107	AR095107 Sequence
12	22	100.0	1186	6	AR098313	AR098313 Sequence
13	22	100.0	1186	6	AX012338	AX012338 Sequence
14	22	100.0	1186	6	I49886	I49886 Sequence 2
15	22	100.0	1186	6	I82374	I82374 Sequence 2
c 16	22	100.0	2476	12	TBI251013	AJ251013 Transform
17	22	100.0	3200	6	I44104	I44104 Sequence 23
18	22	100.0	3201	6	I44103	I44103 Sequence 22
c 19	22	100.0	3236	12	TBI251014	AJ251014 Transform
c 20	22	100.0	4832	6	AX172441	AX172441 Sequence
c 21	22	100.0	4946	6	A60108	A60108 Sequence 1
c 22	22	100.0	4946	6	A76915	A76915 Sequence 1
c 23	22	100.0	4946	6	AR098307	AR098307 Sequence
c 24	22	100.0	4946	6	AX172440	AX172440 Sequence
c 25	22	100.0	5349	6	A71437	A71437 Sequence 7
c 26	22	100.0	5560	6	A60112	A60112 Sequence 5
c 27	22	100.0	5560	6	AR098311	AR098311 Sequence
28	22	100.0	5865	6	AX127748	AX127748 Sequence
c 29	22	100.0	5865	6	AX127748	AX127748 Sequence
30	22	100.0	6539	6	E31991	E31991 Mutated bar
31	22	100.0	6548	6	A60109	A60109 Sequence 2
32	22	100.0	6548	6	A76916	A76916 Sequence 2
33	22	100.0	6548	6	AR098308	AR098308 Sequence
34	22	100.0	6548	6	E31990	E31990 Mutated bar
35	22	100.0	7566	6	A24783	A24783 plasmid pPS
36	22	100.0	7566	6	AR074388	AR074388 Sequence
37	22	100.0	7599	6	AX063413	AX063413 Sequence
38	22	100.0	7639	6	A24782	A24782 plasmid pJD
39	22	100.0	7639	6	AR074387	AR074387 Sequence
c 40	22	100.0	7811	6	AR078675	AR078675 Sequence
41	22	100.0	12095	12	BINHYGDNA	Z37515 Binary vect
c 42	22	100.0	24595	1	ATACH5	X00493 Agrobacteri
c 43	22	100.0	24595	6	E00404	E00404 Ti plasmid
c 44	22	100.0	24595	6	E00546	E00546 DNA fragmen
c 45	22	100.0	194140	1	AF242881	AF242881 Agrobacte

ALIGNMENTS

RESULT 1

AX127753

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

AX127753

Sequence 6 from Patent WO0131042.

AX127753

AX127753.1 GI:14134400

synthetic construct.

synthetic construct

artificial sequence.

1 (bases 1 to 22)

Weston,B. and de Beuckeleer,M.

Male-sterile brassica plants and methods for producing same

Patent: WO 0131042-A 6 03-MAY-2001;

Aventis CropScience N.V. (BE)

Location/Qualifiers

1..22

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer MDB193"

PAT

15-MAY-2001

```

BASE COUNT      6 a      7 c      4 g      5 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22
|||||
Db 1 TCATCTACGGCAATGTACCAGC 22

RESULT 2
AXI72462      AXI72462      22 bp      DNA      PAT      03-JUL-2001
DEFINITION    Sequence 23 from Patent WO0141558.
ACCESSION     AXI72462
VERSION       AXI72462.1 GI:14597574
SOURCE        synthetic construct.
ORGANISM      artificial sequence.
REFERENCE     1 (bases 1 to 22)
AUTHORS       de Both, G. and de Beuckeleer, M.
TITLE         Hybrid winter oilseed rape and methods for producing same
JOURNAL       Patent: WO 0141558-A 23 14-JUN-2001;
              Aventis CropScience N.V. (BE)
FEATURES     Location/Qualifiers
             source
               1..22
               /organism="synthetic construct"
               /db_xref="taxon:32630"
               /note="primer 193"
BASE COUNT      6 a      7 c      4 g      5 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22
|||||
Db 1 TCATCTACGGCAATGTACCAGC 22

RESULT 3
ARGMTUB
LOCUS          249 bp      DNA      SYN      02-APR-1988
DEFINITION     Soybean beta-1-tubulin gene fused to Ti plasmid unit 7 30TR.
ACCESSION      X05579
VERSION        X05579.1 GI:58087
KEYWORDS       beta-tubulin; fusion gene; plasmid.
SOURCE         synthetic construct.
ORGANISM       artificial sequence.
REFERENCE      1 (bases 1 to 249)
AUTHORS        Guiltinan, M.J., Veiten, J., Bustos, M.M., Cyr, R.J., Schell, J. and
              Fosket, D.E.
TITLE          The expression of a chimeric soybean beta-tubulin gene in tobacco
JOURNAL        Mol. Gen. Genet. 207, 328-334 (1987)
FEATURES     Location/Qualifiers
             source
               1..249
               /organism="synthetic construct"
               /db_xref="taxon:32630"
               1..51
               /note="fusion product (17AA); Protein sequence is in
               conflict with the conceptual translation"
               /codon_start=1
               /transl_table=11
               /protein_id="CAA29084.1"
               /db_xref="GI:4376141"
               /translation="MAAAWASSNWSSTDPpMS"
CDS

```

---

```

misc_feature     1..13
                 /note="beta-1-tubulin sequence"
misc_feature     14..28
                 /note="pUC 13 polylinker"
misc_feature     29..35
                 /note="Sal I linker"
misc_feature     37
                 /note="theroretical fusion junction (24) with gene 7 of T1
                 plasmid"
misc_feature     173..178
                 /note="put.polyA signal"
polyA_site       198
                 /note="polyA site"
misc_feature     220..225
                 /note="put.polyA signal"
BASE COUNT      76 a      47 c      38 g      88 t
ORIGIN

Query Match      100.0%; Score 22; DB 12; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22
|||||
Db 106 TCATCTACGGCAATGTACCAGC 127

RESULT 4
ATTRN7
LOCUS          831 bp      DNA      BCT      02-SEP-1999
DEFINITION     Agrobacterium tumefaciens gene encoding 'transcript 7' mRNA for a
              protein with unknown function.
ACCESSION      V00090
VERSION        V00090.1 GI:39180
KEYWORDS       unidentified reading frame.
SOURCE         Agrobacterium tumefaciens.
ORGANISM       Agrobacterium tumefaciens.
              Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
              Rhizobiales; Rhizobium.
REFERENCE      1 (bases 1 to 831)
AUTHORS        Dhaese, P., De Greve, H., Gielen, J., Seurinck, J., Van Montagu, M.M.
              and Schell, J.
TITLE          Identification of sequences involved in the polyadenylation of
              higher plant nuclear transcripts using Agrobacterium T-DNA genes as
              models
JOURNAL        EMBO J. 2, 419-426 (1983)
REFERENCE      2 (bases 76 to 100)
AUTHORS        Dhaese, P.
TITLE          Direct Submission
JOURNAL        Submitted (27-MAY-1983) to the EMBL/GenBank/DBJ databases
COMMENT        Data kindly reviewed (27-MAY-1983) by Dhaese P.
FEATURES     Location/Qualifiers
             source
               1..831
               /organism="Agrobacterium tumefaciens"
               /strain="octopine TL-DNA"
               /db_xref="taxon:358"
               75..81
               /note="CG [1] revised CCAGAGG [2]"
               /citation=1
               /citation=2
               99..101
               /note="CCA [1] revised CTA [2]"
               /citation=1
               /citation=2
               132..673
               /note="transcript 7 (alternate)"
               132..672
               /note="transcript 7"
               148..528
               /note="unknown gene (148 is 1st base in codon) (525 is 3rd
               base in codon)"
               /codon_start=1
old_sequence
old_sequence
mrna
mrna
CDS

```



```
/transl_table=11
/protein_id="CAA23429.1"
/db_xref="GI:39181"
/translation="MNFADTPLASLDLDWACEEFIKTYGASPOLETGEVIOTNNGLLY
LYGKLSORIHDTHLKPKKEELSFTTIKPAEMKAQSDLTYYVAIFQSNVFLCVSN
PEKGFLECHNRPFLYPIVAGSMS"
BASE COUNT      262 a   176 c   138 g   255 t
ORIGIN

Query Match      100.0%; Score 22; DB 1; Length 831;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 580 TCATCTACGGCAATGTACCAGC 601

RESULT 5
ATTDNA      878 bp      DNA      BCT      25-MAR-1996
LOCUS      Agrobacterium tumefaciens crown gall tumor T-DNA from Ti (tumor
DEFINITION      inducing) plasmid pTiA6.
ACCESSION      X00431.1 GI:39150
VERSION      plasmid.
KEYWORDS      Agrobacterium tumefaciens.
SOURCE      Agrobacterium tumefaciens.
ORGANISM      Agrobacterium tumefaciens.
REFERENCE      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
AUTHORS      Rhizobiaceae; Rhizobium.
TITLE      McPherson,J.C.
DNA sequence analysis of crown gall tumor T-DNA encoding the 0.7 kb
transcript
JOURNAL      Nucleic Acids Res. 12 (5), 2317-2325 (1984)
MEDLINE      84169535
FEATURES      Location/Qualifiers
source      1..878
organism="Agrobacterium tumefaciens"
strain="plasmid pTiA6"
db_xref="taxon:358"
60..66
/note="TATA-box"
68..613
/note="polyadenylation signal"
109..489
/note="unidentified reading frame"
/codon_start=1
/transl_table=11
/protein_id="CAA23129.1"
/db_xref="GI:39151"
/db_xref="SWISS-PROT:P03867"
/translation="MNFADTPLASLDLDWACEEFIKTYGASPOLETGEVIOTNNGLLY
LYGKLSORIHDTHLKPKKEELSFTTIKPAEMKAQSDLTYYVAIFQSNVFLCVSN
PEKGFLECHNRPFLYPIVAGSMS"
659..664
/note="polyadenylation signal"
BASE COUNT      288 a   189 c   139 g   262 t
ORIGIN

Query Match      100.0%; Score 22; DB 1; Length 878;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 541 TCATCTACGGCAATGTACCAGC 562

RESULT 6
```

```
A10942
LOCUS      A10942      1037 bp      DNA      PAT      27-SEP-1993
DEFINITION      Nucleotide sequence 4 from patent number DE3920034.
ACCESSION      A10942
VERSION      A10942.1 GI:492369
KEYWORDS      .
SOURCE      unidentified.
ORGANISM      unidentified.
REFERENCE      1 (bases 1 to 1037)
AUTHORS      Patent: DE 3920034-A 4 31-MAY-1990;
JOURNAL      Location/Qualifiers
FEATURES      1..1037
source      /organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      338 a   174 c   166 g   359 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 1037;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 837 TCATCTACGGCAATGTACCAGC 858

RESULT 7
A10939
LOCUS      A10939      1085 bp      DNA      PAT      27-SEP-1993
DEFINITION      Nucleotide sequence 1 from patent number DE3920034.
ACCESSION      A10939
VERSION      A10939.1 GI:492367
KEYWORDS      .
SOURCE      unidentified.
ORGANISM      unidentified.
REFERENCE      1 (bases 1 to 1085)
AUTHORS      Patent: DE 3920034-A 1 31-MAY-1990;
JOURNAL      Location/Qualifiers
FEATURES      1..1085
source      /organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      369 a   218 c   155 g   343 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 1085;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 885 TCATCTACGGCAATGTACCAGC 906

RESULT 8
A10943
LOCUS      A10943      1160 bp      DNA      PAT      27-SEP-1993
DEFINITION      Nucleotide sequence 5 from patent number DE3920034.
ACCESSION      A10943
VERSION      A10943.1 GI:492370
KEYWORDS      .
SOURCE      unidentified.
ORGANISM      unidentified.
REFERENCE      1 (bases 1 to 1160)
AUTHORS      Patent: DE 3920034-A 5 31-MAY-1990;
JOURNAL      Location/Qualifiers
FEATURES
```

```

source
1. .1160
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 367 a 194 c 188 g 411 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1160;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
|||||
Db 960 TCATCTACGGCAATGTACCAGC 981

RESULT 9
AL0941
LOCUS AR0941 1166 bp DNA PAT 27-SEP-1993
DEFINITION Nucleotide sequence 3 from patent number DE3920034.
ACCESSION AR0941
VERSION AR0941.1 GI:492368
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1166)
AUTHORS
JOURNAL Patent: DE 3920034-A 3 31-MAY-1990;
FEATURES
LOCATION/Qualifiers
1. .1166
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 381 a 208 c 196 g 381 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1166;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
|||||
Db 1066 TCATCTACGGCAATGTACCAGC 1087

RESULT 10
AL18051
LOCUS AR18051 1186 bp DNA PAT 26-JUL-1994
DEFINITION DNA used as a probe for neo gene seq ID No:2.
ACCESSION AR18051
VERSION AR18051.1 GI:593120
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1186)
AUTHORS
JOURNAL Patent: WO 9209696-A 2 11-JUN-1992;
FEATURES
LOCATION/Qualifiers
1. .1186
/organism="synthetic construct"
/db_xref="taxon:32630"
1. _8
/promoter
/notes="sequence derived from tapetum specific promoter of
Nicotiana tabacum"
167. .790
gene
/genes="neomycine phosphotransferase gene"
167. .790
CDS
/genes="neomycine phosphotransferase gene"
/notes="Protein sequence is in conflict with the conceptual
translation"
/codon_start=1

/transl_table=11
/protein_id="CAA01373.1"
/db_xref="GI:4529900"
/translation="NELODEAARLSWLATTGVPCAAVLDVVTEAGRDWLLIGVEVPGQD
LSSHLPAAEKVSIAMADMRRLHTLDPATCFDQAKHRIERARTMEAGLVQDDDL
EEHQGLAPAEFLAKRMPDGEDLVVTHGDACLPLNIMVNGRFSGFIDCGRLGVADR
YQDIALATRDIAEELGGEWADRFVLVYGAAPDSQRTAFYRLDDEF"
1055. .1186
/notes="3' regulatory sequence containing the
polyadenylation site derived from agrobacterium T-DNA gene
7"
BASE COUNT 244 a 317 c 325 g 300 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 11
AR095107
LOCUS AR095107 1186 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 2 from patent US 6002070.
ACCESSION AR095107
VERSION AR095107.1 GI:10022665
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1186)
AUTHORS D'Halluin,K. and Gobel,E.
TITLE Process for transforming monocotyledonous plants
JOURNAL Patent: US 6002070-A 2 14-DEC-1999;
FEATURES
LOCATION/Qualifiers
1. .1186
/organism="unknown"
/db_xref="taxon:32630"
BASE COUNT 244 a 317 c 325 g 300 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 12
AR098313
LOCUS AR098313 1186 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 2 from patent US 6074877.
ACCESSION AR098313
VERSION AR098313.1 GI:12807570
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1186)
AUTHORS D'Halluin,K. and Gobel,E.
TITLE Process for transforming monocotyledonous plants
JOURNAL Patent: US 6074877-A 2 13-JUN-2000;
FEATURES
LOCATION/Qualifiers
1. .1186
/organism="unknown"
/db_xref="taxon:32630"
BASE COUNT 244 a 317 c 325 g 300 t
ORIGIN

```

Query Match 100.0%; Score 22; DB 6; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||  
Db 1029 TCATCTACGCAATGTACCAGC 1050

RESULT 13  
AX012338  
LOCUS AX012338 1186 bp DNA PAT 06-SEP-2000  
DEFINITION Sequence 2 from Patent EP0955371.  
ACCESSION AX012338  
VERSION AX012338.1 GI:9998387  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 1186)  
AUTHORS D'Halluin,K. and Goebel,E.D.  
TITLE Process for transforming monocotyledonous plants  
JOURNAL Patent: EP 0955371-A 2 10-NOV-1999;  
PLANT GENETIC SYSTEMS NV (BE)  
FEATURES  
Source 1..1186  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="DNA used as probe for neo gene"

misc\_feature 1..8  
/note="sequence derived from tapetum specific promoter of Nicotiana tabacum"  
misc\_feature 9..790  
/note="coding sequence of neomycine phosphotransferase"  
misc\_feature 791..1186  
/note="3' regulatory sequence containing the polyadenylation site derived from Agrobacterium T-DNA gene"  
BASE COUNT 244 a 317 c 325 g 300 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||  
Db 1029 TCATCTACGCAATGTACCAGC 1050

RESULT 14  
I49886  
LOCUS I49886 1186 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 2 from patent US 5641664.  
ACCESSION I49886  
VERSION I49886.1 GI:2472106  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1186)  
AUTHORS D'Halluin,K. and Goebel,E.  
TITLE Process for transforming monocotyledonous plants  
JOURNAL Patent: US 5641664-A 2 24-JUN-1997;  
FEATURES  
Source 1..1186  
/organism="unknown"  
BASE COUNT 244 a 317 c 325 g 300 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||  
Db 1029 TCATCTACGCAATGTACCAGC 1050

RESULT 15  
I82374  
LOCUS I82374 1186 bp DNA PAT 10-JUN-1998  
DEFINITION Sequence 2 from patent US 5712135.  
ACCESSION I82374  
VERSION I82374.1 GI:3210671  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 1186)  
AUTHORS D'Halluin,K. and Gobel,E.  
TITLE Process for transforming monocotyledonous plants  
JOURNAL Patent: US 5712135-A 2 27-JAN-1998;  
FEATURES  
Source 1..1186  
/organism="unknown"  
BASE COUNT 244 a 317 c 325 g 300 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.76;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||  
Db 1029 TCATCTACGCAATGTACCAGC 1050

Search completed: February 25, 2002, 18:00:10  
Job time: 18418 sec

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:24 ; Search time 716.55 Seconds  
(without alignments)  
26.322 Million cell updates/sec

Title: US-09-698-903B-6

Perfect score: 22

Sequence: 1 tcattctacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

```

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	22	100.0	22	AAH25438	PCR primer for the
2	22	100.0	22	AAH25438	PCR primer MDB193
3	22	100.0	1037	AAH25438	USP-Promoter-casse
4	22	100.0	1085	AAH25438	Legumin-signalpept
5	22	100.0	1160	AAH25438	USP-signalpeptide
6	22	100.0	1166	AAH25438	USP-Promoter-casse
7	22	100.0	1186	AAH25438	Chimeric neo gene
8	22	100.0	1303	AAH25438	Plasmid pTS88 (Eco
9	22	100.0	3153	AAH25438	Plasmid pTS88 (Eco
10	22	100.0	3201	AAH25438	pVS029 Bt ICP codi
11	22	100.0	3201	AAH25438	pVS36 Bt ICP codin

12	22	100.0	3336	21	AAZ29121	Plasmid DV130 comp
13	22	100.0	3694	21	AAZ29124	Plasmid DV133 used
14	22	100.0	3877	21	AAZ29123	Plasmid DV132 used
15	22	100.0	4832	22	AAH25423	Nucleotide sequenc
16	22	100.0	4946	18	AAH25423	T-DNA of plasmid p
17	22	100.0	4946	22	AAH25422	Nucleotide sequenc
18	22	100.0	5228	22	AAH25422	Plasmid pTS172delt
19	22	100.0	5349	19	AAV23239	T-DNA of pTR524.
20	22	100.0	5864	17	AAH23339	Plasmid pTCO113 T-
21	22	100.0	5864	17	AAH23339	Plasmid pTCO113 T-
22	22	100.0	5865	22	AAH23339	Chimeric T-DNA of
23	22	100.0	5865	22	AAH23339	Chimeric T-DNA of
24	22	100.0	6539	21	AAH23339	E. coli plasmid pT
25	22	100.0	6548	17	AAH23336	Plasmid pTS174 use
26	22	100.0	6548	18	AAH23336	Plasmid pTS172. C
27	22	100.0	6548	21	AAH23336	E. coli plasmid pT
28	22	100.0	7492	22	AAH23336	Plasmid pTS346. U
29	22	100.0	7566	14	AAH23336	Plasmid pTS0212 co
30	22	100.0	7599	22	AAH23320	Nucleotide sequenc
31	22	100.0	7639	14	AAH23320	Plasmid pTD884 con
32	22	100.0	24593	6	AAH50226	Sequence of opine
33	22	100.0	24596	6	AAH50226	Complete nucleotid
34	21	95.5	21	22	AAH89355	T-DNA right border
35	18	81.8	19	20	AAH28157	Transformed Arabid
36	18	81.8	20	22	AAH28157	PCR primer for the
37	18	81.8	20	22	AAH28157	PCR primer MDB258
38	18	81.8	24	21	AAH28157	PCR primer T4 used
39	18	81.8	415	22	AAH28157	Right (5') border
40	18	81.8	416	22	AAH28157	Left (3') border f
41	17.4	79.1	1077	22	AAH28157	Right flanking reg
42	17.2	78.2	1565	21	AAH28157	Fusarium venenatum
43	16.4	74.5	2558	13	AAH28157	Human liver GPI-PL
44	16.4	74.5	2558	16	AAH28157	Glycosyl-phosphati
45	16.4	74.5	14244	20	AAH28157	Polynucleotide seq

#### ALIGNMENTS

RESULT 1	
AAH25438	
ID	AAH25438 standard; DNA; 22 BP.
XX	
AC	AAH25438;
XX	
DT	22-AUG-2001 (first entry)
XX	
DE	PCR primer for the right flanking region in transgenic plant BN-RF1.
XX	
KW	Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW	fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.
XX	
OS	Synthetic.
XX	
PN	WO200141558-A1.
XX	
PD	14-JUN-2001.
XX	
PF	06-DEC-2000; 2000WO-EF12872.
XX	
PR	08-DEC-1999; 99US-0457037.
XX	
PA	(AVET ) AVENTIS CROPS SCIENCE NV.
XX	
PI	De Both G, De Beuckeleer M;
XX	
DR	WPI; 2001-381419/40.
XX	
PT	Transgenic winter oilseed rape plants suited for producing hybrid seed
PT	with improved qualities, comprises a male-sterility gene and fertility
PT	restorer gene, integrated into the genome
XX	
PS	Claim 60; Page 46; 98pp; English.

CC The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC The present PCR primer was used to amplify the right flanking region of  
 CC a vector in a transgenic plant which carries the TA29-barstar  
 CC transgene.

SQ Sequence 22 BP; 6 A; 7 C; 4 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.034;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 ||||||||||||||||||  
 Db 1 tcattctacggcaatgtaccagc 22

## RESULT 2

AA006995  
 ID AAD06995 standard; DNA; 22 BP.

AC AAD06995;

DT 06-AUG-2001 (first entry)

DE PCR primer M09193 to generate the flanking region of elite event MS-B2.

KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; thermal asymmetric interlaced;  
 KW TAIL; ss.

OS Agrobacterium sp.

PN WO200131042-A2.

PD 03-MAY-2001.

PF 26-OCT-2000; 2000WO-EP10680.

PR 29-OCT-1999; 99US-0430497.

PA (AVET ) AVENTIS CROPS SCIENCE NV.

PI Weston B, De Beuckeleer M;

DR WPI; 2001-300517/31.

PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -

PS Example 3; Page 28; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is secondary thermal interlaced (TAIL)-PCR primer  
 CC M09193 used to right (5') border flanking region of elite event MS-B2.  
 CC This primer corresponds to position 226-247 of plasmid pTC0113.

XX

SQ Sequence 22 BP; 6 A; 7 C; 4 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.034;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 ||||||||||||||||||  
 Db 1 tcattctacggcaatgtaccagc 22

## RESULT 3

AA004705  
 ID AA004705 standard; DNA; 1037 BP.

AC AA004705;

DT 12-OCT-1990 (first entry)

DE USP-Promoter-cassette USP-Pr.T7.1.

KW Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.

PN DE3920034-A.

PD 31-MAY-1990.

PF 20-JUN-1989; 89DE-3920034.

PR 19-SEP-1988; 88DD-0319887.

PA (PFLA-) VE KOMB PFLANZENZUC.

PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;  
 DR WPI; 1990-172459/23.

PT Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.

PS Disclosure; ; pp; German.

CC The unique BglII-Ort (720-725) site is for ligating foreign DNA and the  
 CC HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 CC cassette in the T1-vector pGA471. Agrobacterium tumefaciens is  
 CC transfected.

CC See also AA004703-Q04706.

XX SQ Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 100.0%; Score 22; DB 11; Length 1037;  
 Best Local Similarity 100.0%; Pred. No. 0.059;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 ||||||||||||||||||  
 Db 837 tcattctacggcaatgtaccagc 858

## RESULT 4

AA004703  
 ID AA004703 standard; DNA; 1085 BP.

AC AA004703;

DT 12-OCT-1990 (first entry)

DE Legumin-signalpeptide cassette Le-Sig.T7.

XX Foreign DNA incorporation; recombinant DNA techniques;

KW higher plant genome; legumin; B4 gene; signalpeptide; Le-Sig.T7.; ss.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 747..814  
 FT /\*tag=a  
 FT /product=Legumin-signalpeptide  
 XX  
 PN DE3920034-A.  
 XX  
 PD 31-MAY-1990.  
 XX  
 PF 20-JUN-1989; 89DE-3920034.  
 XX  
 PR 19-SEP-1988; 88DD-0319887.  
 XX  
 XX (PFLA-) VE KOMB PFLANZENZUC.  
 XX  
 PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;  
 XX  
 DR WPI; 1990-172459/23.  
 DR P-PSDB; AAR05198.  
 XX  
 PT Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.  
 XX  
 PS Disclosure; ; pp; German.  
 XX  
 CC The Legumin gene B4 is used. The unique BglII-Ort (815-820) site is for  
 CC ligating foreign DNA and the HindIII-Ort in the 3' polylinker  
 CC (1080-1085) for cloning the cassette in the Ti-vector pGA471.  
 CC The cassette is cloned into the binary Ti-vectors pGA471 and  
 CC Agrobacterium tumefaciens is transfected.  
 CC See also AAQ04703-Q04706.  
 XX  
 SQ Sequence 1085 BP; 368 A; 219 C; 155 G; 343 T; 0 other;  
 XX  
 Query Match 100.0%; Score 22; DB 11; Length 1085;  
 Best Local Similarity 100.0%; Pred. No. 0.059;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcattacgcgcaatgtaccagc 22  
 Db 885 tcattacgcgcaatgtaccagc 906  
 RESULT 5  
 AAQ04706  
 ID AAQ04706 standard; DNA; 1160 BP.  
 XX  
 AC AAQ04706;  
 XX  
 DT 12-OCT-1990 (first entry)  
 XX  
 DE USP-signalpeptide cassette USP-Sig.T7.  
 XX  
 KW Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; signalpeptide; USP-Sig.T7.; ss.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 708..877  
 FT /\*tag=a  
 FT /product=signalpeptide  
 FT intron 747..817  
 XX  
 PN DE3920034-A.  
 XX  
 PD 31-MAY-1990.  
 XX  
 PF 20-JUN-1989; 89DE-3920034.  
 XX  
 PR 19-SEP-1988; 88DD-0319887.  
 XX

PA (PFLA-) VE KOMB PFLANZENZUC.  
 XX  
 PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;  
 XX  
 DR WPI; 1990-172459/23.  
 DR P-PSDB; AAR05199.  
 XX  
 PT Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.  
 XX  
 PS Disclosure; ; pp; German.  
 XX  
 CC The unique BglII-Ort (890-895) site is for  
 CC ligating foreign DNA and the HindIII-Ort in the 3' polylinker  
 CC (1155-1160) for cloning the cassette in the Ti-vector pGA471.  
 CC The cassette is cloned into the binary Ti-vectors pGA471 and  
 CC Agrobacterium tumefaciens is transfected.  
 CC See also AAQ04703-Q04706.  
 XX  
 SQ Sequence 1160 BP; 369 A; 192 C; 188 G; 411 T; 0 other;  
 XX  
 Query Match 100.0%; Score 22; DB 11; Length 1160;  
 Best Local Similarity 100.0%; Pred. No. 0.06;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcattacgcgcaatgtaccagc 22  
 Db 960 tcattacgcgcaatgtaccagc 981  
 RESULT 6  
 AAQ04704  
 ID AAQ04704 standard; DNA; 1166 BP.  
 XX  
 AC AAQ04704;  
 XX  
 DT 12-OCT-1990 (first entry)  
 XX  
 DE USP-Promoter-cassette USP-Pr.T7.2.  
 XX  
 KW Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-2; ss.  
 XX  
 PN DE3920034-A.  
 XX  
 PD 31-MAY-1990.  
 XX  
 PF 20-JUN-1989; 89DE-3920034.  
 XX  
 PR 19-SEP-1988; 88DD-0319887.  
 XX  
 XX (PFLA-) VE KOMB PFLANZENZUC.  
 PA  
 PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;  
 XX  
 DR WPI; 1990-172459/23.  
 XX  
 PT Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.  
 XX  
 PS Disclosure; ; pp; German.  
 XX  
 CC The unique BglII-Ort (720-725) site is for ligating foreign DNA and the  
 CC HindIII-Ort in the 3' polylinker (1261-1266) for cloning the  
 CC cassette in the Ti-vector pGA471. Agrobacterium tumefaciens is  
 CC transfected.  
 CC See also AAQ04703-Q04706.  
 XX  
 SQ Sequence 1166 BP; 381 A; 208 C; 196 G; 381 T; 0 other;  
 XX  
 Query Match 100.0%; Score 22; DB 11; Length 1166;

```

Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcaccggaatgtaccagc 22
    |||||
Db 1066 tcattcaccggaatgtaccagc 1087

RESULT 7
AAQ25707
ID AAQ25707 standard; DNA; 1186 BP.
XX
AC AAQ25707;
XX
DT 07-DEC-1992 (first entry)
XX
DE Chimeric neo gene probe.
XX
KW Nicotiana tabacum; neomycine phosphotransferase; Agrobacterium; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT promoter 1..8
FT /tag= a
FT /note= "sequence derived from tapetum specific
FT promoter of Nicotiana tabacum"
FT CDS 9..790
FT /tag= b
FT /product= neomycine_phosphotransferase
FT 791..1186
FT /tag= c
FT /note= "3', regulatory sequence contg. the
FT polyadenylation site derived from
FT Agrobacterium T-DNA gene 7"
XX
PN W09209696-A.
XX
PD 11-JUN-1992.
XX
PF 21-NOV-1991; 91WO-EP02198.
XX
PR 23-NOV-1990; 90EP-0403332.
PR 08-JUL-1991; 91EP-0401888.
XX
PA (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
PI Dhalluin K, Goebel E;
XX
DR WPI; 1992-217075/26.
XX
PT Transforming monocotyledonous plants e.g. cereals - comprises
PT wounding and/or degrading cells of intact plant tissue or
PT embryogenic callus
XX
PS Disclosure; Page 60; 76pp; English.
XX
CC Two transformed corn plants were analysed by means of Southern
CC hybridisation. As a probe, a 1184 bp EcoRI-HindIII fragment derived
CC from another plasmid was used. The sequence of that plasmid is
CC given below. Results showed that at least a chimeric neo gene was
CC integrated into the plant genomic DNA.
XX
SQ Sequence 1186 BP; 244 A; 317 C; 325 G; 300 T; 0 other;

Query Match 100.0%; Score 22; DB 13; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcaccggaatgtaccagc 22
    |||||
Db 1029 tcattcaccggaatgtaccagc 1050

```

```

RESULT 8
AAT39337
ID AAT39337 standard; DNA; 1303 BP.
XX
AC AAT39337;
XX
DT 22-JAN-1997 (first entry)
XX
DE Plasmid pTS88 (EcoRI-HindIII fragment).
XX
KW Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1..35
FT /tag= a
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
FT promoter 36..694
FT /tag= b
FT /label= P35S
FT /function= 35S promoter of cauliflower mosaic virus
FT strain CM1841
FT CDS 695..967
FT /tag= c
FT /label= barstar
FT /product= Bacillus amyloliquefaciens barstar
FT 968..1287
FT /tag= d
FT /label= 3'g7
FT /function= region containing polyadenylation signal
FT of gene 7 og Agrobacterium T-DNA
FT misc_feature 1288..1303
FT /tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX
PN W09626283-A1.
XX
PD 29-AUG-1996.
XX
PF 21-FEB-1996; 96WO-EP00722.
XX
PR 21-FEB-1995; 95EP-0400364.
XX
PA (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
PI Botterman J, Cornelissen M, Michiels F;
XX
DR WPI; 1996-402373/40.
XX
PT Prodn. of male sterile plants by transforming with a chimeric
PT construct - comprising a male sterility DNA e.g. barnase and a
PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
PT for generating hybrid cultivars
XX
PS Example 1; Page 38; 56pp; English.
XX
CC The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains
CC barstar DNA under control of a 35S promoter. The plasmid was
CC used with pTS174 (see also AAT39336) contg. barnase DNA under
CC control of the stamen-specific promoter Ei to produce male sterile
CC rice cv. Koshihikari transgenic plants, and with plasmid pV8136
CC (see also AAT39336) contg. barnase DNA under control of the stamen-
CC specific pCA55 promoter to produce male sterile maize plants.
CC Expression of barnase (a ribonuclease) in the stamen leads to male
CC sterility. Constitutive expression of barstar counteracts possible
CC low level expression of barnase DNA in non-stamen tissue.
XX

```



SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 100.0%; Score 22; DB 17; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 0.061;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||

Db 1094 tcattacggcaatgtaccagc 1115  
|||||

## RESULT 9

AAZ29122  
ID AAZ29122 standard; DNA; 3153 BP.

XX AC AAZ29122;

XX DT 21-FEB-2000 (first entry)

XX DE Plasmid DV131 comprising L3/EGFP:NPTII/Tr7 expression cassette.

XX KW Transgenic seed; marker: aleurone-specific promoter; Plasmid DV131;  
KW GFP:NPTII fusion protein construct; L3/EGFP:NPTII/Tr7; maize R gene;  
KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene; assay;  
KW Green fluorescent protein; GFP; gene fusion; selection; screening;  
KW expression; automated seed screening technique; screenable marker;  
KW transformant; embryogenic tissue; implementation; ds.

XX OS Synthetic.

XX PN WO9960129-A1.

XX PD 25-NOV-1999.

XX PF 18-MAY-1999; 99WO-US11023.

XX PR 18-MAY-1998; 98US-0080625.

XX PA (DEKA-) DEKALB GENETICS CORP.

XX PI Kriz AL, Spencer TW;

XX DR WPI; 2000-072441/06.

XX PT Screenable marker genes useful for identification of transgenic seeds  
XX for plant breeding -

XX PS Example 1; Page 164-166; 182pp; English.

XX CC The present DNA sequence is the plasmid DV131, that is used in the  
CC generation of GFP:NPTII fusion protein constructs. It contains an  
CC expression cassette comprising, a promoter from the maize L3 oleosin  
CC gene, the coding sequence of EGFP:NPTII translational fusion, excised  
CC from DV126 and the Tr7 terminator. This plasmid is used to carry a  
CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
CC operably linked to the gene encoding a screenable marker, like Green  
CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
CC allow both selection and screening of transformants. The aleurone-  
CC specific promoters direct the expression of the marker genes in  
CC embryogenic tissues, allowing selection and screening of viable  
CC transgenic seeds. Screening of transgenic seeds avoids the need for  
CC growing and assaying of seeds for transgenes and allows implementation  
CC of automated seed screening techniques for the identification of  
XX transgenic seeds.

XX SQ Sequence 3153 BP; 801 A; 857 C; 756 G; 739 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3153;  
Best Local Similarity 100.0%; Pred. No. 0.069;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||

Db 2675 tcattacggcaatgtaccagc 2696  
|||||

## RESULT 10

AAQ14529  
ID AAQ14529 standard; DNA; 3201 BP.

XX AC AAQ14529;

XX DT 27-JAN-1992 (first entry)

XX DE pPS029 Bt ICP coding sequence.

XX KW Bacillus thuringiensis; insecticidal crystal protein; ICP;  
KW deletion; ss.

XX OS Synthetic.

XX PN WO9116432-A.

XX PD 31-OCT-1991.

XX PF 17-APR-1991; 91WO-EP00733.

XX PR 18-APR-1990; 90EP-0401055.

XX PA (PLAN-) PLANT GENETIC SYST.

XX PI Cornelissen M, Soetaert P, Stam M, Dockx J;

XX DR WPI; 1991-339820/46.

XX PT Modified Bacillus thuringiensis insecticidal crystal protein  
XX genes - having A and T sequences changed to G and C sequences  
XX encoding same amino acids, for increased expression levels

XX PS Disclosure; Fig 6(C); 78pp; English.

XX CC "n" in the sequence refers to not known nucleotides.

XX CC pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-  
XX terminal modification and the internal modification of the Bt ICP  
XX coding sequence.

XX CC See also AAQ14529, AAQ15142-44.

XX SQ Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;

## Query Match

Best Local Similarity 100.0%; Score 22; DB 12; Length 3201;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||

Db 2948 tcattacggcaatgtaccagc 2969  
|||||

## RESULT 11

AAQ15144  
ID AAQ15144 standard; DNA; 3201 BP.

XX AC AAQ15144;

XX DT 27-JAN-1992 (first entry)

XX DE pVE36 Bt ICP coding sequence.

XX KW Bacillus thuringiensis; insecticidal crystal protein; ICP;  
KW deletion; ss.

XX OS Synthetic.

```

XX PN WO9116432-A.
XX PD 31-OCT-1991.
XX PF 17-APR-1991; 91WO-EP00733.
XX PR 18-APR-1990; 90EP-0401055.
XX PA (PLAN-) PLANT GENETIC SYST.
XX PI Cornelissen M, Soetaert P, Stam M, Dockx J;
XX PF WI; 1991-339820/46.
XX PT Modified Bacillus thuringiensis insecticidal crystal protein
PT genes - having A and T sequences changed to G and C sequences
PT encoding same amino acids, for increased expression levels
XX PS Disclosure; Fig 6(C); 78pp; English.
XX CC "n" in the sequence refers to not known nucleotides.
CC PPS029 (AAQ14529) is identical to pVE36, but carries both the amino-
CC terminal modification and the internal modification of the Bt ICP
CC coding sequence.
CC See also AAQ14529, AAQ15142-44.
XX SQ Sequence 3201 BP; 915 A; 651 C; 700 G; 930 T; 5 other;

Query Match 100.0%; Score 22; DB 12; Length 3201;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcgaatgtaccagc 22
Db 3021 tcattctacggcgaatgtaccagc 3042

RESULT 12
AAZ29121
ID AAZ29121 standard; DNA; 3336 BP.
XX AC AAZ29121;
XX DT 21-FEB-2000 (first entry)
XX DE Plasmid DV130 comprising L3/MGFP:NPTII/Tr7 expression cassette.
XX KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DV130;
KW GFP:NPTII fusion protein construct; L3/MGFP:NPTII/Tr7; maize R gene;
KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;
KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;
KW Green fluorescent protein; GFP; gene fusion; selection; screening;
KW transformant; expression; embryogenic tissue; implementation; ds.
XX OS Synthetic.
XX PN WO9960129-A1.
XX PD 25-NOV-1999.
XX PF 18-MAY-1999; 99WO-US11023.
XX PR 18-MAY-1998; 98US-0080625.
XX PA (DEKA-) DEKALB GENETICS CORP.
XX PI Kriz AL, Spencer TM;
XX DR WI; 2000-072441/06.
XX PT Screenable marker genes useful for identification of transgenic seeds
XX for plant breeding -
XX Example 1; Page 168-170; 182pp; English.
XX CC The present DNA sequence is the plasmid DV133, comprising

```

```

PT Screenable marker genes useful for identification of transgenic seeds
PT for plant breeding -
XX Example 1; Page 163-164; 182pp; English.
XX CC The present DNA sequence is the plasmid DV130, that is used in the
CC generation of GFP:NPTII fusion protein constructs. It contains an
CC expression cassette comprising, a promoter from the maize L3 oleosin
CC gene, the coding sequence of MGFP:NPTII translational fusion, excised
CC from DV127 and the Tr7 terminator. This plasmid is used to carry a
CC chimeric gene, comprising an aleurone-specific promoter like L3, that is
CC operably linked to a gene encoding a screenable marker, like Green
CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions
CC allow both selection and screening of transformants. The aleurone-
CC specific promoters direct the expression of the marker genes in
CC embryogenic tissues, allowing selection and screening of viable
CC transgenic seeds. Screening of transgenic seeds avoids the need for
CC growing and assaying of seeds for transgenes and allows implementation
CC of automated seed screening techniques for the identification of
XX transgenic seeds.
XX SQ Sequence 3336 BP; 888 A; 830 C; 764 G; 854 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3336;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcgaatgtaccagc 22
Db 2858 tcattctacggcgaatgtaccagc 2879

RESULT 13
AAZ29124
ID AAZ29124 standard; DNA; 3694 BP.
XX AC AAZ29124;
XX DT 21-FEB-2000 (first entry)
XX DE Plasmid DV133 used for construction of GFP:NPTII fusion protein.
XX KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DV133;
KW GFP:NPTII fusion protein construct; L3/Tr7 intron/EGFP:NPTII/Tr7;
KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;
KW EGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;
KW Green fluorescent protein; GFP; gene fusion; maize R gene; selection;
KW screenable marker; screening; GFP; automated seed screening technique; assay;
KW transformant; expression; embryogenic tissue; implementation; ds.
XX OS Synthetic.
XX PN WO9960129-A1.
XX PD 25-NOV-1999.
XX PF 18-MAY-1999; 99WO-US11023.
XX PR 18-MAY-1998; 98US-0080625.
XX PA (DEKA-) DEKALB GENETICS CORP.
XX PI Kriz AL, Spencer TM;
XX DR WI; 2000-072441/06.
XX PT Screenable marker genes useful for identification of transgenic seeds
XX for plant breeding -
XX Example 1; Page 168-170; 182pp; English.
XX CC The present DNA sequence is the plasmid DV133, comprising

```

CC L3/rACT11 Intron/EGFP:NPTII/Tr7, that is used in the generation of  
 CC GFP:NPTII fusion protein constructs. It contains an expression cassette  
 CC comprising, a promoter from the maize L3 oleosin gene, the rice actin 1  
 CC intron, the coding sequence of EGFP:NPTII translational fusion, excised  
 CC from DV126 and the Tr7 terminator. This plasmid is used to carry a  
 CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to a gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.

XX  
 SQ Sequence 3694 BP; 890 A; 984 C; 914 G; 906 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3694;  
 Best Local Similarity 100.0%; Pred. No. 0.07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcgaatgtaccagc 22  
 |||||||||||||||||||  
 Db 3216 tcattctacggcgaatgtaccagc 3237

RESULT 14  
 AAZ29123  
 ID AAZ29123 standard; DNA; 3877 BP.  
 XX  
 AC AAZ29123;  
 XX  
 DT 21-FEB-2000 (first entry)  
 XX  
 DE Plasmid DV132 used for construction of GFP:NPTII fusion protein.  
 XX  
 KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DV132;  
 KW GFP:NPTII fusion protein construct; L3/rACT11 intron/MGFP:NPTII/Tr7;  
 KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
 KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;  
 KW Green fluorescent protein; GFP; gene fusion; maize R gene; selection;  
 KW screenable marker; screening; automated seed screening technique; assay;  
 KW transformant; expression; embryogenic tissue; implementation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9960129-A1.  
 XX  
 PD 25-NOV-1999.  
 XX  
 PF 18-MAY-1999; 99WO-US11023.  
 XX  
 PR 18-MAY-1998; 98US-0080625.  
 XX  
 PA (DEKA-) DEKALB GENETICS CORP.  
 XX  
 PI Kriz AL, Spencer TM;  
 XX  
 DR WPI; 2000-072441/06.  
 XX  
 XX Screenable marker genes useful for identification of transgenic seeds  
 PT for plant breeding -  
 PT  
 XX Example 1; Page 166-168; 182pp; English.  
 PS  
 XX The present DNA sequence is the plasmid DV132, comprising  
 CC L3/rACT11 Intron/MGFP:NPTII/Tr7, that is used in the generation of  
 CC GFP:NPTII fusion protein constructs. It contains an expression cassette  
 CC comprising, a promoter from the maize L3 oleosin gene, the rice actin 1  
 CC intron, the coding sequence of MGFP:NPTII translational fusion, excised  
 CC from DV127 and the Tr7 terminator. This plasmid is used to carry a

CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to a gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.

XX  
 SQ Sequence 3877 BP; 977 A; 957 C; 922 G; 1021 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3877;  
 Best Local Similarity 100.0%; Pred. No. 0.071;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcgaatgtaccagc 22  
 |||||||||||||||||||  
 Db 3399 tcattctacggcgaatgtaccagc 3420

RESULT 15  
 AAH25423/C  
 ID AAH25423 standard; DNA; 4832 BP.  
 XX  
 AC AAH25423;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 DE Nucleotide sequence of plasmid pTHW118.  
 XX  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barstar gene; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 OS Arabidopsis thaliana.  
 OS Bacillus amyloliquefaciens.  
 OS Nicotiana tabacum.  
 XX

Key Location/Qualifiers  
 FT misc\_feature 1..25 /tag= a  
 FT /note= "Right border repeat from TL-DNA from pTiB6S3"  
 FT misc\_feature 26..53 /tag= b  
 FT /note= "synthetic polylinker derived sequences"  
 FT misc\_feature 54..90 /tag= c  
 FT /note= "residual sequence from TL-DNA at right  
 FT border repeat"  
 FT misc\_feature 91..97 /tag= d  
 FT /note= "synthetic polylinker derived sequences"  
 FT 3'UTR complement (98..305)  
 FT /tag= e  
 FT /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"  
 FT misc\_feature 310..330 /tag= f  
 FT /note= "synthetic polylinker derived sequences"  
 FT CDS complement (331..882)  
 FT /tag= g  
 FT /note= "Streptomyces hygroscopicus bialaphos  
 FT resistance (883..2608)  
 FT complement (883..2608)  
 FT /tag= h  
 FT /note= "AtSIA ribulose-1,5-biphosphate carboxylase  
 FT small subunit gene from Arabidopsis thaliana"  
 FT misc\_feature 2609..2658  
 FT /tag= i  
 FT /note= "synthetic polylinker derived sequences"  
 FT

Search completed: February 25, 2002, 18:17:24  
Job time: 16682 sec

```

FT 3'UTR complement (2659..2919)
FT /*tag= j
FT /note= "taql fragment from 3' UTR of nopaline
FT synthase gene from T-DNA of pTiT37 and
FT containing plant polyadenylation signals"
FT misc_feature 2920..2940
FT /*tag= k
FT /note= "synthetic polylinker derived sequences"
FT 3'UTR 2941..2980
FT /*tag= l
FT /note= "downstream of Bacillus amyloliquefaciens
FT barstar coding region"
FT CDS complement (2981..3253)
FT /*tag= m
FT /note= "Barstar gene coding region from Bacillus
FT amyloliquefaciens"
FT promoter complement (3254..4762)
FT /*tag= n
FT /note= "anther-specific gene TA29 promoter from
FT Nicotiana tabacum"
FT misc_feature 4763..4807
FT /*tag= o
FT /note= "synthetic polylinker derived sequences"
FT misc_feature 4808..4832
FT /*tag= p
FT /note= "left border repeat from Tl-DNA from pTiB6S3"
XX WO200141558-A1.
PN 14-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-EPI2872.
XX
XX 08-DEC-1999; 99US-0457037.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX De Both G, De Beuckeleer M;
XX
XX WPI; 2001-381419/40.
XX
XX Transgenic winter oilseed rape plants suited for producing hybrid seed
XX with improved qualities, comprises a male-sterility gene and fertility
XX restorer gene, integrated into the genome
XX
XX Example 1; Page 80-82; 98pp; English.
XX
XX The specification describes a pair of transgenic winter oilseed rape
XX plants suited for producing hybrid seed. One of the plants has an
XX expression cassette comprising a male-sterility gene, and the other
XX plant has an expression cassette comprising a fertility restorer gene,
XX integrated into the genome. The fertility restorer gene is capable of
XX preventing the activity of the male-sterility gene. The plant pair is
XX useful for producing hybrid seed. Plants developed from the hybrid
XX seed have agronomic performance, genetic stability and adaptability to
XX different genetic backgrounds. The present sequence represents
XX plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
XX a fertility restorer gene. The plasmid is used to create transgenic
XX plants of the invention.
XX
XX Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;
XX
XX Query Match 100.0%; Score 22; DB 22; Length 4832;
XX Best Local Similarity 100.0%; Pred. No. 0.073;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 tcattctacggcaatgtaccagc 22
XX |
XX Db 247 TCATCTACGGCAATGTACCAGC 226

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:15 ; Search time 301.6 Seconds  
(without alignments)  
16.520 Million cell updates/sec

Title: US-09-698-903B-6  
Perfect score: 22  
Sequence: 1 tcattctacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues  
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	1186	1 US-08-064-121-2	Sequence 2, Appli
2	22	100.0	1186	1 US-08-478-015-2	Sequence 2, Appli
3	22	100.0	1186	3 US-08-475-975-2	Sequence 2, Appli
4	22	100.0	1186	3 US-09-084-889-2	Sequence 2, Appli
5	22	100.0	1303	3 US-08-894-440-2	Sequence 2, Appli
6	22	100.0	3153	4 US-09-080-625-3	Sequence 3, Appli
7	22	100.0	3200	1 US-08-453-104-23	Sequence 23, Appli
8	22	100.0	3200	2 US-08-694-824-23	Sequence 23, Appli
9	22	100.0	3201	1 US-08-453-104-22	Sequence 22, Appli
10	22	100.0	3201	2 US-08-694-824-22	Sequence 22, Appli
11	22	100.0	3336	4 US-09-080-625-5	Sequence 2, Appli
12	22	100.0	3694	4 US-09-080-625-5	Sequence 5, Appli
13	22	100.0	3877	4 US-09-080-625-4	Sequence 4, Appli
c 14	22	100.0	4946	3 US-08-817-188-1	Sequence 1, Appli
c 15	22	100.0	5560	3 US-08-817-188-5	Sequence 5, Appli
c 16	22	100.0	5864	3 US-08-894-440-4	Sequence 4, Appli
c 17	22	100.0	5864	3 US-08-894-440-4	Sequence 4, Appli
18	22	100.0	6548	3 US-08-894-440-1	Sequence 1, Appli
19	22	100.0	6548	3 US-08-817-188-2	Sequence 2, Appli
20	22	100.0	7566	2 US-08-232-016-23	Sequence 23, Appli
21	22	100.0	7639	2 US-08-232-016-22	Sequence 22, Appli
c 22	22	100.0	7811	2 US-08-549-680A-5	Sequence 5, Appli
c 23	18.8	85.5	24595	6 5428147-1	Patent No. 5428147
c 24	16	72.7	29	2 US-08-232-016-12	Sequence 12, Appli
c 25	15.8	71.8	984	4 US-09-446-504-4	Sequence 4, Appli
c 26	15.8	71.8	3574	4 US-09-446-504-83	Sequence 83, Appli
c 27	15.8	71.8	3620	4 US-09-446-504-55	Sequence 55, Appli

c 28	15.2	69.1	238	1 US-07-914-284A-1	Sequence 1, Appli
c 29	15.2	69.1	238	5 PCT-US93-06645-1	Sequence 1, Appli
c 30	14.8	67.3	1290	4 US-09-247-373B-55	Sequence 55, Appli
c 31	14.8	67.3	1347	1 US-08-663-713A-1	Sequence 1, Appli
c 32	14.8	67.3	1347	3 US-09-014-888-1	Sequence 1, Appli
c 33	14.6	66.4	617	2 US-08-392-546C-3	Sequence 3, Appli
34	14.6	66.4	1512	4 US-08-955-918C-8	Sequence 8, Appli
35	14.6	66.4	1801	4 US-08-955-918C-6	Sequence 6, Appli
36	14.6	66.4	1804	1 US-08-631-200-1	Sequence 1, Appli
37	14.6	66.4	1804	1 US-08-829-553-1	Sequence 1, Appli
38	14.6	66.4	1804	2 US-08-922-267A-1	Sequence 1, Appli
39	14.6	66.4	1804	2 US-08-936-707A-1	Sequence 1, Appli
40	14.6	66.4	1804	2 US-08-936-706A-1	Sequence 1, Appli
41	14.6	66.4	1804	3 US-09-248-203-1	Sequence 1, Appli
42	14.6	66.4	1804	4 US-09-406-071-1	Sequence 1, Appli
43	14.6	66.4	2119	1 US-08-630-592-1	Sequence 1, Appli
44	14.6	66.4	2119	1 US-08-714-991-1	Sequence 1, Appli
45	14.6	66.4	2119	3 US-09-032-365A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-064-121-2  
; Sequence 2, Application US/08064121  
; Patent No. 5641664  
; GENERAL INFORMATION:  
; APPLICANT: D'HALLUIN, Kathleen  
; APPLICANT: GOBEL, Elke  
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; Zip: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/064,121  
; FILING DATE: 24-MAY-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 90403332.1  
; FILING DATE: 23-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91401888.2  
; FILING DATE: 08-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crane-Feury, Sharon E  
; REGISTRATION NUMBER: 36,113  
; REFERENCE/DOCKET NUMBER: 010830-043  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1186 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: probe

```
;
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
;
US-08-064-121-2
```

```
Query Match 100.0%; Score 22; DB 1; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050
```

```
RESULT 2
US-08-478-015-2
; Sequence 2, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2;
```

```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine phosphotransferase ge
;
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylat
; OTHER INFORMATION: site derived from Agrobacterium T-DNA gene 7"
;
US-08-478-015-2

Query Match 100.0%; Score 22; DB 1; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 3
US-08-475-975-2
; Sequence 2, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-08-475-975-2
;
; Query Match 100.0%; Score 22; DB 3; Length 1186;
; Best Local Similarity 100.0%; Pred. No. 0.02;
; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 tcattctacggcaatgtaccagc 22
; |||||
; DB 1029 TCATCTACGGCAATGTACCAGC 1050
;
; RESULT 4
; US-09-084-889-2
; Sequence 2, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/064,121
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-09-084-889-2
;
; Query Match 100.0%; Score 22; DB 3; Length 1186;
; Best Local Similarity 100.0%; Pred. No. 0.02;
; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 tcattctacggcaatgtaccagc 22
; |||||
; DB 1029 TCATCTACGGCAATGTACCAGC 1050
;
; RESULT 5
; US-08-894-440-2
; Sequence 2, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NWSGOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
; OTHER INFORMATION: fragment of pTS88
```

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(35)
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (35)..(694)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain
; OTHER INFORMATION: CM1841 (P35S)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (695)..(967)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (968)..(1287)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1288)..(1303)
; OTHER INFORMATION: polylinker of pGEM2
;
; US-08-894-440-2
;
; Query Match 100.0%; Score 22; DB 3; Length 1303;
; Best Local Similarity 100.0%; Pred. No. 0.02;
; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 tcattacggcaatgtaccagc 22
;      |||||
; Db 1094 tcattacggcaatgtaccagc 1115
```

```
RESULT 6
US-09-080-625-3
; Sequence 3, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; APPLICANT: Spencer, T. Michael
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
;
; TOPOLOGY: linear
; US-09-080-625-3
;
; Query Match 100.0%; Score 22; DB 4; Length 3153;
; Best Local Similarity 100.0%; Pred. No. 0.024;
; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 tcattacggcaatgtaccagc 22
;      |||||
; Db 2675 TCATCTACGGCAATGTACCAGC 2696
;
; RESULT 7
; US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,104
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note="Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
;
; US-08-453-104-23
;
; Query Match 100.0%; Score 22; DB 1; Length 3200;
; Best Local Similarity 100.0%; Pred. No. 0.024;
; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



QY 1 tcattacggcaatgtaccagc 22  
|||||

Db 2948 TCATCTACGGCAATGTACCAGC 2969

## RESULT 8

US-08-694-824-23  
; Sequence 23, Application US/08694824

; Patent No. 5877306

; GENERAL INFORMATION:

; APPLICANT: CORNELISSEN, Marc

; APPLICANT: SOETAERT, Piet

; APPLICANT: STAM, Maïke

; APPLICANT: DOCKX, Jan

; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS

; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/694,824

; FILING DATE: 09-AUG-1996

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/937,869

; FILING DATE: 16-DEC-1992

; APPLICATION NUMBER: GB 90401055.0

; FILING DATE: 18-APR-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Rea, Teresa S

; REGISTRATION NUMBER: 30,427

; REFERENCE/DOCKET NUMBER: 010830-032

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 23:

; LENGTH: 3200 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: 2078..2082

; OTHER INFORMATION: /note= "Nucleotides 2078-2082

; OTHER INFORMATION: wherein N is not known."

US-08-694-824-23

Query Match 100.0%; Score 22; DB 2; Length 3200;

Best Local Similarity 100.0%; Pred. No. 0.024;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22

|||||

Db 2948 TCATCTACGGCAATGTACCAGC 2969

## RESULT 9

US-08-453-104-22

; Sequence 22, Application US/08453104

; Patent No. 5633446

; GENERAL INFORMATION:

; APPLICANT: CORNELISSEN, Marc

; APPLICANT: SOETAERT, Piet

; APPLICANT: STAM, Maïke

; APPLICANT: DOCKX, Jan

; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS

; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/453,104

; FILING DATE:

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/937,869

; FILING DATE: 16-DEC-1992

; APPLICATION NUMBER: GB 90401055.0

; FILING DATE: 18-APR-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Rea, Teresa S

; REGISTRATION NUMBER: 30,427

; REFERENCE/DOCKET NUMBER: 010830-032

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3201 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: 2151..2155

; OTHER INFORMATION: /note= "Nucleotides 2151-2155

; OTHER INFORMATION: wherein N is not known."

US-08-453-104-22

Query Match 100.0%; Score 22; DB 1; Length 3201;

Best Local Similarity 100.0%; Pred. No. 0.024;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22

|||||

Db 3021 TCATCTACGGCAATGTACCAGC 3042

## RESULT 10

US-08-694-824-22

; Sequence 22, Application US/08694824

; Patent No. 5877306

; GENERAL INFORMATION:

; APPLICANT: CORNELISSEN, Marc

; APPLICANT: SOETAERT, Piet

; APPLICANT: STAM, Maïke

; APPLICANT: DOCKX, Jan

; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS

; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

```
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,824
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 18-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 2151..2155
; OTHER INFORMATION: /note= "Nucleotides 2151-2155
; OTHER INFORMATION: wherein N is not known."
US-08-694-824-22

Query Match 100.0%; Score 22; DB 2; Length 3201;
Best Local Similarity 100.0%; Pred. NO. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22
|||||
Db 3021 TCATCTACGCGCAATGTACCAGC 3042

RESULT 11
US-09-080-625-2
; Sequence 2, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3336 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-080-625-2

Query Match 100.0%; Score 22; DB 4; Length 3336;
Best Local Similarity 100.0%; Pred. NO. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22
|||||
Db 2858 TCATCTACGCGCAATGTACCAGC 2879

RESULT 12
US-09-080-625-5
; Sequence 5, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-080-625-5
```

Query Match 100.0%; Score 22; DB 4; Length 3694;  
 Best Local Similarity 100.0%; Pred. No. 0.024; 0; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 |||||  
 DB 3216 TCATCTACGGCAATGTACCAGC 3237

RESULT 13  
 US-09-080-625-4  
 ; Sequence 4, Application US/09080625  
 ; Patent No. 6307123  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kriz, Alan L.  
 ; APPLICANT: Spencer, T. Michael  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
 ; TITLE OF INVENTION: IDENTIFICATION  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold White & Durkee  
 ; STREET: P.O. Box 4433  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77210-4433  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/080,625  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hanson, Robert E.  
 ; REGISTRATION NUMBER: P-42,628  
 ; REFERENCE/DOCKET NUMBER: DEKM:161  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (512) 418-3000  
 ; TELEFAX: (512) 474-7577  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3877 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-09-080-625-4

Query Match 100.0%; Score 22; DB 4; Length 3877;  
 Best Local Similarity 100.0%; Pred. No. 0.024;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 |||||  
 DB 3399 TCATCTACGGCAATGTACCAGC 3420

RESULT 14  
 US-08-817-188-1/c  
 ; Sequence 1, Application US/08817188  
 ; Patent No. 6074876  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DE BLOCK, MARC  
 ; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 ; FILE REFERENCE: 2121-0127P  
 ; CURRENT APPLICATION NUMBER: US/08/817,188  
 ; CURRENT FILING DATE: 1997-05-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
 ; EARLIER FILING DATE: 1996-07-31

; EARLIER APPLICATION NUMBER: EP 95401844.6  
 ; EARLIER FILING DATE: 1995-08-04  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 4946  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
 ; OTHER INFORMATION: plasmid pTHW107  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement(1)..(25)  
 ; OTHER INFORMATION: T-DNA right border (RB)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement(97)..(330)  
 ; OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
 ; OTHER INFORMATION: T-DNA  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement(331)..(882)  
 ; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
 ; OTHER INFORMATION: transferase  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((883)..(2608))  
 ; OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
 ; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((2658)..(3031))  
 ; OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
 ; OTHER INFORMATION: gene of Agrobacterium T-DNA  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((3032)..(3367))  
 ; OTHER INFORMATION: barnase: region coding for barnase  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((3368)..(4876))  
 ; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana  
 ; OTHER INFORMATION: tabacum  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((4922)..(4946))  
 ; OTHER INFORMATION: LB: T-DNA left border  
 ; US-08-817-188-1

Query Match 100.0%; Score 22; DB 3; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.025;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 |||||  
 DB 247 TCATCTACGGCAATGTACCAGC 226

RESULT 15  
 US-08-817-188-5/c  
 ; Sequence 5, Application US/08817188  
 ; Patent No. 6074876  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DE BLOCK, MARC  
 ; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 ; FILE REFERENCE: 2121-0127P  
 ; CURRENT APPLICATION NUMBER: US/08/817,188  
 ; CURRENT FILING DATE: 1997-05-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/03366

```
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTI1B6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-LS1 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5087)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTI1B6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
```

US-08-817-188-5

```
Query Match      100.0%; Score 22; DB 3; Length 5560;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
   |||||
Db 234 TCATCTACGGCAATGTACCAGC 213
```

Search completed: February 25, 2002, 18:05:16  
Job time: 18564 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:01 ; Search time 8261.74 Seconds  
(without alignments)  
28.615 Million cell updates/sec

Title: US-09-698-903B-6  
Perfect score: 22  
Sequence: 1 tcctctacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					
c 1	17.4	79.1	568	11	B1174209
2	17.2	78.2	330	10	AW358852
3	17.2	78.2	352	10	AA475114
4	17.2	78.2	355	10	BE032643
5	17.2	78.2	355	11	BE924637
6	17.2	78.2	432	10	AA212655
7	17.2	78.2	562	13	AQ724420
8	17.2	78.2	567	13	AZ851447
9	17.2	78.2	721	11	BG592665
10	17.2	78.2	721	11	BG964419
c 11	17.2	78.2	932	11	BG400800
12	16.8	76.4	162	13	AZ474239

13	16.8	76.4	855	11	BG669143
c 14	16.4	74.5	479	10	AA458389
c 15	16.2	73.6	216	10	BE168752
c 16	16.2	73.6	309	13	TA105D12Q
17	16.2	73.6	314	13	AZ319004
c 18	16.2	73.6	343	10	AW789094
c 19	16.2	73.6	352	11	T69520
c 20	16.2	73.6	413	10	AA637160
21	16.2	73.6	423	10	AI593113
22	16.2	73.6	428	10	AI280109
c 23	16.2	73.6	435	11	T88735
24	16.2	73.6	438	10	AV668989
25	16.2	73.6	440	11	N62954
26	16.2	73.6	443	10	AA878226
c 27	16.2	73.6	450	10	AA256224
28	16.2	73.6	450	10	AA574351
c 29	16.2	73.6	457	11	BG730070
c 30	16.2	73.6	495	10	AI449899
31	16.2	73.6	501	10	AI338197
c 32	16.2	73.6	502	11	BI234655
c 33	16.2	73.6	505	13	AZ720352
c 34	16.2	73.6	508	10	AI862985
c 35	16.2	73.6	544	10	AI533411
36	16.2	73.6	553	10	AI580125
c 37	16.2	73.6	571	10	BE367030
c 38	16.2	73.6	609	10	BE367017
c 39	16.2	73.6	611	11	W37106
c 40	16.2	73.6	622	11	BE819452
c 41	16.2	73.6	644	11	BG049365
c 42	16.2	73.6	658	10	AW953929
43	16.2	73.6	701	10	BE542678
44	16.2	73.6	739	11	BF177958
45	16.2	73.6	753	10	BE393004

ALIGNMENTS

RESULT	1
LOCUS	B1174209/c
DEFINITION	OSTF013F8_1 AD-wrmcDNA Caenorhabditis elegans cDNA similar to C46F11.2, mRNA sequence.
ACCESSION	B1174209
VERSION	B1174209.1
KEYWORDS	EST.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans.
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS	Reboul,J., Vaglio,P., Tzellas,N., Thierry-Mieg,N., Moore,T., Jackson,C., Shin-i,T., Kohara,Y., Thierry-Mieg,D., Thierry-Mieg,J., Lee,H., Hittl,J., Doucette-Stamm,L., Hartley,J.L., Temple,G.F., Brasch,M.A., Vandenhaute,J., Lamesch,P.E., Hill,D.E. and Vidal,M.
TITLE	Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in C. elegans
JOURNAL	Nat. Genet. 27 (3), 332-336 (2001)
MEDLINE	21135099
COMMENT	Contact: Reboul J, Vaglio P Marc Vidal Laboratory Dana Farber Cancer Institute 44 Binney Street, Boston, MA 02115, USA Tel: 617 632 5180 Fax: 617 632 2425 Email: Jerome.Reboul@dfci.harvard.edu Sequence tag of Gateway entry clones. The primers used were designed on the predicted protein encoding ORF. C. elegans ORFeome cloning project; Contact jerome_reboul@dfci.harvard.edu or Philippe_vaglio@dfci.harvard.edu POLYA-No.
FEATURES	Location/Qualifiers
source	1..568

/organism="Caenorhabditis elegans"  
 /strain="N2"  
 /db\_xref="taxon:6239"  
 /clone\_lib="AD-wrmcDNA"  
 /sex="Hermaphrodite and male"  
 /tissue\_type="whole animal"  
 /dev\_stage="mixed stage"  
 /note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPCR86"

BASE COUNT 146 a 116 c 159 g 147 t

ORIGIN

Query Match 79.1%; Score 17.4; DB 11; Length 568;  
 Best Local Similarity 94.7%; Pred. No. 3.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tctacggcaatgtaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 564 TCTACGGCAATATACAGC 546

RESULT 2  
 AW358852  
 LOCUS AW358852 330 bp mRNA EST 09-JUL-2000  
 DEFINITION 43779 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION AW358852  
 VERSION AW358852.1 GI:6863502  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 1 (bases 1 to 330)  
 Fahnenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keeler, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 20 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCCAGTCACGACG  
 Plate: 25 row: A column: 16  
 Seq primer: ATTAGTGACACTAG.

FEATURES  
 source  
 1..330  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MARC 2P1G"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 54 a 118 c 108 g 50 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 10; Length 330;  
 Best Local Similarity 86.4%; Pred. No. 4.2e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcatctacggcaatgtaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 190 TCATCCACGGCGTTGTACCAGC 211

RESULT 3  
 AA475114  
 LOCUS AA475114 352 bp mRNA EST 18-JUN-1997

DEFINITION v095a03.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA  
 clone IMAGE:873676 5', mRNA sequence.

ACCESSION AA475114  
 VERSION AA475114.1 GI:2199644  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 352)

TITLE The WashU-HHMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:513156

Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 346.

FEATURES  
 source  
 1..352  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_image="873676"  
 /clone\_lib="Barstead mouse pooled organs MPLRB4"  
 /sex="mixed"  
 /tissue\_type="pooled organs"  
 /dev\_stage="7 day"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTTGGATCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bob Barstead."

BASE COUNT 57 a 118 c 101 g 75 t  
 ORIGIN

Query Match 78.2%; Score 17.2; DB 10; Length 352;  
 Best Local Similarity 86.4%; Pred. No. 4.2e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcatctacggcaatgtaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 54 TCATCTACCGGATTGTACCAGC 75

RESULT 4  
 BE032643





polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCGAATGGAGCCGCCGGAATCTTTTITTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 71 a 148 c 125 g 88 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 10; Length 432;  
Best Local Similarity 86.4%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
||||||| | | | | | | | | | |

Db 94 TCATCTACCGGATGTACCAGC 115

RESULT 7  
A0724420  
LOCUS  
DEFINITION HS\_2119\_A2\_C07\_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2119 Col-14 Row=E, DNA sequence.  
ACCESSION A0724420  
VERSION A0724420.1 GI:5484089  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 562)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 2119 row: E column: 14  
Seq primer: M13 Reverse  
Class: BAC ends

High quality sequence stop: 562.

FEATURES  
source  
Location/Qualifiers  
1..562  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 142 a 137 c 122 g 153 t 8 others

ORIGIN

Query Match 78.2%; Score 17.2; DB 13; Length 562;  
Best Local Similarity 86.4%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
||||||| | | | | | | | | | |

Db 438 TCATCTACGCCACTGTGCCAGC 459

RESULT 8  
LOCUS  
DEFINITION

AZ851447 567 bp DNA GSS 21-FEB-2001  
2M0153M11R Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGC2M0153M11 R, DNA sequence.

ACCESSION AZ851447  
VERSION AZ851447.1 GI:13037453  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 567)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00  
Plate: 0153 row: M column: 11

Seq primer: CACACAGGAACAGCATGACCC

Class: plasmid ends

High quality sequence stop: 567.

FEATURES  
Location/Qualifiers  
1..567

source

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0153M11"  
/clone\_lib="Mouse 10kb plasmid UUGCLM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil473211419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 188 a 94 c 106 g 179 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 13; Length 567;  
Best Local Similarity 86.4%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22

Email: cgabbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10996 row: m column: 16  
High quality sequence stop: 718.

```
FEATURES
source
1..721
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4986807"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      124 a   223 c   228 g   146 t
ORIGIN
Query Match          78.2%; Score 17.2; DB 11; Length 721;
Best Local Similarity 86.4%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 tcattctacggcaatgtaccagc 22  
||||| ||||| ||||| |||||  
Db 611 TCATCTACCGCATGTACCAGC 632

```
RESULT 11
BG400800/c
LOCUS       BG400800             932 bp     mRNA           EST       12-MAR-2001
DEFINITION BG2464016Fl NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592474 5',
            mRNA sequence.
ACCESSION   BG400800
VERSION     BG400800.1 GI:13294248
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 932)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
UNPUBLISHED (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            CDNA Library Preparation: CLONTECH Laboratories, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1330 row: o column: 03
            High quality sequence stop: 230.
```

```
FEATURES
source
1..932
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4592474"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcccatgcc); Site_2: SfiI (ggccatgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
```

---

Db 300 TCAACTACAGCAATGAACCAGC 279  
||||| ||||| ||||| |||||

```
RESULT 9
BG592665      721 bp     mRNA           EST       12-APR-2001
LOCUS       EST491343 cSTS Solanum tuberosum cDNA clone cSTS2C3 5' sequence,
            mRNA sequence.
DEFINITION   BG592665
ACCESSION   BG592665
VERSION     BG592665.1 GI:13610805
KEYWORDS    EST.
SOURCE      Potato.
ORGANISM     Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1 (bases 1 to 721)
AUTHORS     van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE       Generations of ESTs from sprouting potato eyes
JOURNAL     Unpublished (2000)
COMMENT     Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: MI3F-R.
            Location/Qualifiers
            1..721
            /organism="Solanum tuberosum"
            /cultivar="Kennebec"
            /db_xref="taxon:4113"
            /clone="cSTS2C3"
            /clone_lib="cSTS"
            /tissue_type="sprouting eyes from tubers"
            /dev_stage="12-14 weeks post harvest"
            /lab_host="SOLR"
            XhoI; Various sizes of sprouting eyes (2mm to 15mm) were taken from tubers. The tubers were incubated at 26C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers."
```

```
BASE COUNT      192 a   202 c   113 g   214 t
ORIGIN
Query Match          78.2%; Score 17.2; DB 11; Length 721;
Best Local Similarity 86.4%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 tcattctacggcaatgtaccagc 22  
||||| ||||| ||||| |||||  
Db 282 TCGTCCACCGCAATGTACCAGC 303

```
RESULT 10
BG964419      721 bp     mRNA           EST       12-JUN-2001
LOCUS       BG2832063Fl NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4986807 5',
            mRNA sequence.
DEFINITION   BG964419
ACCESSION   BG964419
VERSION     BG964419.1 GI:14352056
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 721)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```



QY 3 atctacggcaatgtaccagc 22  
||||| ||||||| |||||  
Db 673 ATCTAGGCAATGTCCACG 692

RESULT 14  
AA458389/c  
LOCUS 479 bp mRNA EST 06-JUN-1997  
DEFINITION vq49f10.r1 Soares mammary\_gland\_NbMMG Mus musculus cDNA clone  
IMAGE:864715 5' similar to gb:J03040 SPARC PRECURSOR (HUMAN);  
gb:M20692 Mouse osteonectin (MOUSE);, mRNA sequence.  
ACCESSION AA458389  
VERSION AA458389.1 GI:2181109  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 479)  
Maira.M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Thelasing,B., Wyllie,T., Lennon,G., Soares,B., Wilsson,R. and  
Waterston,R.  
TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:508803  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 441.

FEATURES  
source  
1. 479  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:864715"  
/clone\_lib="Soares\_mammary\_gland\_NbMMG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia  
RI: 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5',  
TGTTCACCAATCTGAAGTGGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldo."  
BASE COUNT 114 a 125 c 127 g 112 t 1 others  
ORIGIN

Query Match 74.5%; Score 16.4; DB 10; Length 479;  
Best Local Similarity 94.4%; Pred. NO. 1.1e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 tcattctacggcaatgtac 18  
||||| ||||||| |||||  
Db 312 TCATCCACGGCAATGTAC 295

RESULT 15  
BE168752/c  
LOCUS 216 bp mRNA EST 21-JUN-2000  
DEFINITION QVI-HT0516-140300-107-e07 HT0516 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE168752  
VERSION BE168752.1 GI:8631473  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 216)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=st2=QV1-HT0516-140  
300-107-e07&t3=2000-03-14&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 216.  
FEATURES  
source  
1. 216  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0516"  
/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 42 a 75 c 59 g 40 t  
ORIGIN  
Query Match 73.6%; Score 16.2; DB 10; Length 216;  
Best Local Similarity 85.7%; Pred. NO. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 tcattctacggcaatgtaccag 21  
||||| ||||||| |||||  
Db 46 TCAGCCACGGCAAGTACCAG 26

Search completed: February 25, 2002, 17:21:04  
Job time: 16157 sec

---







```
RESULT 4
AX063413/c
LOCUS AX063413 7599 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 5 from Patent WO0100833.
ACCESSION AX063413
VERSION AX063413.1 GI:12541201
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
1 (bases 1 to 7599)
REFERENCE
AUTHORS Hofmann,B., Mollier,P. and Pelletier,G.
TITLE Promoter expressed specifically in the cells of plant roots,
recombinant vectors and host cells comprising same and transgenic
plants obtained
JOURNAL Patent: WO 0100833-A 5 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES
source
Location/Qualifiers
1..7599
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="ADN-T de PKB5"
BASE COUNT 1972 a 1938 c 1937 g 1752 t
ORIGIN
Query Match 46.1%; Score 191.4; DB 6; Length 7599;
Best Local Similarity 99.5%; Pred. No. 5.6e-25;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 223 gatgtacatgcccgaataaggaagcaattttagatgttaattcccatcttgaagaaa 282
Db 6771 GATGTACATGTCGATAGAAAGGCAATTTCTAGATGTTAATTCCTTCCATCTTGAAGAAA 6712
QY 283 tatagttaaattattattgataaaatacaaaagtcaggtattattagtcacaaagcaaaaaca 342
Db 6711 TATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACA 6652
QY 343 taaattattgatgcaagtttaaaattcagaataatttcaataaactgattatcagctgg 402
Db 6651 TAAATTTATTGATGCAAGTTTAAATTTCAGAAATATTTTCAATAACTGATTATATCAGCTGG 6592
QY 403 tacattgccgtag 415
Db 6591 TACATTGCCGTAG 6579
RESULT 5
E31991/c
LOCUS E31991 6539 bp DNA PAT 07-FEB-2001
DEFINITION Mutated barnase gene and transgenic plant thereof.
ACCESSION E31991
VERSION E31991.1 GI:13021588
KEYWORDS JP 2000041682-A/4.
SOURCE unidentified.
ORGANISM unclassified.
1 (bases 1 to 6539)
REFERENCE
AUTHORS Kazuyuki,H.F.N.N.
TITLE Mutated barnase gene and transgenic plant thereof
JOURNAL Patent: JP 2000041682-A 4 15-FEB-2000;
JAPAN TOBACCO INC
OS Escherichia coli LE392
PN JP 2000041682-A/4
PD 15-FEB-2000
PF 04-AUG-1998 JP 1998220060
PR
PI KAZUYUKI HAMADA,FUMIO NAKAKIDO
PC C12N15/09,A01H5/00,C12N5/10,C12N9/22//C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00,(C12N5/00,C12R1:91)
CC
```

```

FH Key Location/Qualifiers
FT source 1..6539 /organism='Escherichia coli LE392'.
FEATURES
source
Location/Qualifiers
1..6539
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1755 a 1578 c 1519 g 1687 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6539;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatgcccgaataaggaagcaattttagatgttaattcccatcttgaagaaaat 283
Db 6533 ACGTACATGTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTCCATCTTGAAGAAAAT 6474
QY 284 atagttaaattatttattgataaaatacaaaagtcaggtattattagtcacaaagcaaaaacat 343
Db 6473 ATAGTTTAAATATTTATTCGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACAT 6414
QY 344 aaatttattgatgcaagtttaaaattcagaataatttcaataaactgattatcagctggt 403
Db 6413 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTTCAATAACTGATTATATCAGCTGGT 6354
QY 404 acattgccgtag 415
Db 6353 ACATTGCCGTAG 6342
RESULT 6
A60109/c
LOCUS A60109 6548 bp DNA circular PAT 22-OCT-1999
DEFINITION Sequence 2 from Patent WO9706267.
ACCESSION A60109
VERSION A60109.1 GI:3715125
KEYWORDS
SOURCE Plasmid PTS172.
ORGANISM Plasmid PTS172
REFERENCE 1 (bases 1 to 6548)
AUTHORS De,B.M.
TITLE Genetic transformation using a PAPP inhibitor
JOURNAL Patent: WO 9706267-A 2 20-FEB-1997;
PLANT GENETIC SYSTEMS NV (BE)
FEATURES
source
Location/Qualifiers
1..6548
/organism="Plasmid PTS172"
/db_xref="taxon:106340"
BASE COUNT 1756 a 1579 c 1523 g 1690 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatgcccgaataaggaagcaattttagatgttaattcccatcttgaagaaaat 283
Db 6542 ACGTACATGTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTCCATCTTGAAGAAAAT 6483
QY 284 atagttaaattatttattgataaaatacaaaagtcaggtattattagtcacaaagcaaaaacat 343
Db 6482 ATAGTTTAAATATTTATTCGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACAT 6423
QY 344 aaatttattgatgcaagtttaaaattcagaataatttcaataaactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTTCAATAACTGATTATATCAGCTGGT 6363
QY 404 acattgccgtag 415
Db 6363 ACATTGCCGTAG 6342
```

```

Db 6362 ACATTGCCGTAG 6351

RESULT 7
LOCUS A76916 6548 bp DNA circular PAT 22-OCT-1999
DEFINITION Sequence 2 from Patent EP0757102.
ACCESSION A76916
VERSION A76916.1 GI:6088713
KEYWORDS
SOURCE
ORGANISM
Plasmid PTS172.
Plasmid PTS172.
plasmids.
1 (bases 1 to 6548)
REFERENCE
AUTHORS De.B.M.
TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR
JOURNAL Patent: EP 0757102-A 2 05-FEB-1997;
PLANT GENETIC SYSTEMS NV (BE)
FEATURES
Location/Qualifiers
Source 1..6548
/organism="plasmid PTS172"
/db_xref="taxon:106340"
BASE COUNT 1756 a 1579 c 1523 g 1690 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatgcccgcataagaagaaggcaattttagatgtttaattcccatcttgaaagaacat 283
Db 6542 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATTCATCTTGAAGAAAT 6483

QY 284 atagtttaaatattattgataaaatacaagtcagggtattatagtcacagcaaaacat 343
Db 6482 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACAT 6423

QY 344 aaattttatgatgcaagtttaaatcagaataatttcaataactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATCATCAGCTGGT 6363

QY 404 acattgccgctag 415
Db 6362 ACATTGCCGTAG 6351

RESULT 8
LOCUS AR098308/c
DEFINITION Sequence 2 from patent US 6074876.
ACCESSION AR098308
VERSION AR098308.1 GI:12807565
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 6548)
AUTHORS De Block,M.
TITLE Genetic transformation using a PARP inhibitor
JOURNAL Patent: US 6074876-A 2 13-JUN-2000;
FEATURES
Location/Qualifiers
Source 1..6548
/organism="unknown"
BASE COUNT 1756 a 1579 c 1523 g 1690 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatgcccgcataagaagaaggcaattttagatgtttaattcccatcttgaaagaacat 283
Db 6542 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATTCATCTTGAAGAAAT 6483

QY 284 atagtttaaatattattgataaaatacaagtcagggtattatagtcacagcaaaacat 343
Db 6482 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACAT 6423

QY 344 aaattttatgatgcaagtttaaatcagaataatttcaataactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATCATCAGCTGGT 6363

QY 404 acattgccgctag 415
Db 6362 ACATTGCCGTAG 6351

RESULT 9
LOCUS E31990 6548 bp DNA PAT 07-FEB-2001
DEFINITION Mutated barnase gene and transgenic plant thereof.
ACCESSION E31990
VERSION E31990.1 GI:13021587
KEYWORDS JP 2000041682-A/3.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 6548)
AUTHORS Kazuyuki,H.F.N.N.
TITLE Mutated barnase gene and transgenic plant thereof
JOURNAL Patent: JP 2000041682-A 3 15-FEB-2000;
JAPAN TOBACCO INC
COMMENT OS Escherichia coli LE392
PN JP 2000041682-A/3
PD 15-FEB-2000
PF 04-AUG-1998 JP 1998220060
PR KAZUYUKI HAMADA,FUMIO NAKAKIDO
PC C12N15/09,A01H5/00,C12N5/10,C12N9/22/(C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00,(C12N5/00,C12R1:91)
CC
FH Key Location/Qualifiers
FT source 1..6548
/organism="Escherichia coli LE392".
FEATURES
Location/Qualifiers
Source 1..6548
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1756 a 1579 c 1523 g 1690 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatgcccgcataagaagaaggcaattttagatgtttaattcccatcttgaaagaacat 283
Db 6542 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATTCATCTTGAAGAAAT 6483

QY 284 atagtttaaatattattgataaaatacaagtcagggtattatagtcacagcaaaacat 343
Db 6482 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACAT 6423

QY 344 aaattttatgatgcaagtttaaatcagaataatttcaataactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATCATCAGCTGGT 6363

QY 404 acattgccgctag 415
Db 6362 ACATTGCCGTAG 6351

```

```
RESULT 10
LOCUS AR078675 7811 bp DNA PAT 31-AUG-2000
DEFINITION Sequence 5 from patent US 5962768.
ACCESSION AR078675
VERSION AR078675.1 GI:10005421
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 7811)
AUTHORS Cornelissen,M., Reynaerts,A., Gossele,V. and Van Aarssen,R.
TITLE Marker gene
JOURNAL Patent: US 5962768-A 5 05-OCT-1999;
FEATURES
source
BASE COUNT 1950 a 1906 c 1873 g 2082 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 7811;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatggccgataagaaggaattgttagatgttaattcccatcttgaagaaat 283
DB 438 ACGTACATGGTCGATAGAAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAAT 497
QY 284 atagtttaaatattattatgataaaaatacaagtcagggtatttagtcccaagcaaaaacat 343
DB 498 ATAGTTTAAATATTATTTCATATAATAAACAAGTCAGGTATTATAGTCCCAAGCAAAAACAT 557
QY 344 aaattattgatgcaggtttaattcagaataatttcaataactgaactgattatcagctgct 403
DB 558 AAATTTATTGATCGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGCT 617
QY 404 acattgcgctag 415
DB 618 ACATTGCCGTAG 629
Query Match 45.4%; Score 188.4; DB 6; Length 4832;
Best Local Similarity 99.5%; Pred. No. 2e-24;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
RESULT 11
144104/c
LOCUS I44104 3200 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 23 from patent US 5633446.
ACCESSION I44104
VERSION I44104.1 GI:2469202
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 3200)
AUTHORS Cornelissen,M., Soetaert,P., Stam,M. and Dockx,J.
TITLE Modified Bacillus thuringiensis insecticidal-crystal protein genes
and their expression in plant cells
JOURNAL Patent: US 5633446-A 23 27-MAY-1997;
FEATURES
source
BASE COUNT 880 a 710 c 720 g 885 t 5 others
ORIGIN
Query Match 45.4%; Score 188.4; DB 6; Length 3200;
Best Local Similarity 99.5%; Pred. No. 2e-24;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 226 gtacatggccgataagaaggaattgttagatgttaattcccatcttgaagaaat 285
DB 3141 GTACATGGTCGATAGAAGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 3082
QY 286 agtttaaatattattgataaaatacaagtcagggtatttagtcccaagcaaaaacataa 345
```

```
DB 3081 AGTTTAAATATTATTGTATAAAATACAAGTCAGGTATTATATGTCACAAACAAACATAA 3022
QY 346 atttatgtgcaagtttaaatcagaaataattcaataactgattatcatcagctggtac 405
DB 3021 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTTCANTACTGATTATATCAGCTGCTAC 2962
QY 406 attgcgctag 415
DB 2961 ATTGCCGTAG 2952
RESULT 12
LOCUS AX172441 4832 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 2 from Patent WO0141558.
ACCESSION AX172441
VERSION AX172441.1 GI:14597553
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4832)
AUTHORS de Both,G. and de Beuckeleer,M.
TITLE Hybrid winter oilseed rape and methods for producing same
JOURNAL Patent: WO 0141558-A 2 14-JUN-2001;
FEATURES
source
misc_feature 1528 a 883 c 932 g 1488 t 1 others
BASE COUNT
ORIGIN
Query Match 45.4%; Score 188.4; DB 6; Length 4832;
Best Local Similarity 99.5%; Pred. No. 2e-24;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 226 gtacatggccgataagaaggaattgttagatgttaattcccatcttgaagaaat 285
DB 54 GTACATGGTCGATAGAAGCAATTTGTAGATGTTAATTCCTCATCTTGAAGAAATAT 113
QY 286 agtttaaatattattgataaaatacaagtcagggtatttagtcccaagcaaaaacataa 345
DB 114 AGTTTAAATATTATTGTATAAAATACAAGTCAGGTATTATAGTCCCAAGCAAAAACATAA 173
QY 346 atttatgtgcaagtttaaatcagaaataattcaataactgattatcatcagctggtac 405
DB 174 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTTCANTACTGATTATATCAGCTGCTAC 233
QY 406 attgcgctag 415
DB 234 ATTGCCGTAG 243
RESULT 13
LOCUS A60108 4946 bp DNA PAT 06-MAR-1998
DEFINITION Sequence 1 from Patent WO9706267.
ACCESSION A60108
VERSION A60108.1 GI:3715124
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4946)
AUTHORS De.B.M.
TITLE Genetic transformation using a PARP inhibitor
JOURNAL Patent: WO 9706267-A 1 20-FEB-1997;
```

FEATURES PLANT GENETIC SYSTEMS NV (BE)  
 source Location/Qualifiers  
 1. .4946

BASE COUNT 1569 a 891 c 963 g 1523 t  
 ORIGIN /organism="Transformation vector pTHW107"  
 /db\_xref="taxon:126810"

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
 Best Local Similarity 99.5%; Pred. No. 2e-24;  
 Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgaataagaaagcaattgttagatgtttaattcccatcttgaagaataat 285  
 |||||  
 Db 54 GTACATGGTCGATAGAAAAGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 113  
 |||||  
 QY 286 agtttaaatattattgataaaataacaagtcagggtattattatagtcacaagcaaaaacataa 345  
 |||||  
 Db 114 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACATAA 173  
 |||||  
 QY 346 attattgatgcgaagtttaaatcagaataattccaactgattatcagctggtac 405  
 |||||  
 Db 174 ATTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGGTAC 233  
 |||||  
 QY 406 attgcccgtag 415  
 |||||  
 Db 234 ATTGCCGTAG 243

RESULT 14  
 LOCUS A76915 A76915 4946 bp DNA PAT 19-OCT-1999  
 DEFINITION Sequence 1 from Patent EP0757102.  
 ACCESSION A76915  
 VERSION A76915.1 GI:6088712  
 KEYWORDS Transformation vector pTHW107.  
 SOURCE Transformation vector pTHW107.  
 ORGANISM Transformation vector pTHW107.  
 REFERENCE 1 (bases 1 to 4946)  
 AUTHORS De.B.M.  
 TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 JOURNAL Patent: EP 0757102-A 1 05-FEB-1997;  
 PLANT GENETIC SYSTEMS NV (BE)  
 FEATURES Location/Qualifiers  
 source 1. .4946  
 BASE COUNT 1569 a 891 c 963 g 1523 t  
 ORIGIN /organism="Transformation vector pTHW107"  
 /db\_xref="taxon:126810"

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
 Best Local Similarity 99.5%; Pred. No. 2e-24;  
 Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgaataagaaagcaattgttagatgtttaattcccatcttgaagaataat 285  
 |||||  
 Db 54 GTACATGGTCGATAGAAAAGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 113  
 |||||  
 QY 286 agtttaaatattattgataaaataacaagtcagggtattattatagtcacaagcaaaaacataa 345  
 |||||  
 Db 114 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACATAA 173  
 |||||  
 QY 346 attattgatgcgaagtttaaatcagaataattccaactgattatcagctggtac 405  
 |||||  
 Db 174 ATTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGGTAC 233  
 |||||  
 QY 406 attgcccgtag 415  
 |||||  
 Db 234 ATTGCCGTAG 243

RESULT 15  
 LOCUS AR098307 AR098307 4946 bp DNA PAT 14-FEB-2001  
 DEFINITION Sequence 1 from patent US 6074876.  
 ACCESSION AR098307  
 VERSION AR098307.1 GI:12807564  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 4946)  
 AUTHORS De Block,M.  
 TITLE Genetic transformation using a PARP inhibitor  
 JOURNAL Patent: US 6074876-A 1 13-JUN-2000;  
 FEATURES Location/Qualifiers  
 source 1. .4946  
 BASE COUNT 1569 a 891 c 963 g 1523 t  
 ORIGIN

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
 Best Local Similarity 99.5%; Pred. No. 2e-24;  
 Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgaataagaaagcaattgttagatgtttaattcccatcttgaagaataat 285  
 |||||  
 Db 54 GTACATGGTCGATAGAAAAGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 113  
 |||||  
 QY 286 agtttaaatattattgataaaataacaagtcagggtattattatagtcacaagcaaaaacataa 345  
 |||||  
 Db 114 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACATAA 173  
 |||||  
 QY 346 attattgatgcgaagtttaaatcagaataattccaactgattatcagctggtac 405  
 |||||  
 Db 174 ATTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGGTAC 233  
 |||||  
 QY 406 attgcccgtag 415  
 |||||  
 Db 234 ATTGCCGTAG 243

Search completed: February 25, 2002, 18:00:20  
 Job time: 18428 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:24 ; Search time 716.55 seconds  
(without alignments)  
496.532 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagttgtgttcattga.....cagctggtacattgcgtag 415

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	415	100.0	415	Right (5') border
2	380.8	91.8	416	Left (3') border
3	191.4	46.1	5864	Plasmid pCOLI3 T-
4	191.4	46.1	5865	Chimeric T-DNA of
5	191.4	46.1	7599	Nucleotide sequenc
6	188.8	45.5	5228	Plasmid pTS172del
7	188.8	45.5	6539	E. coli plasmid pT
8	188.8	45.5	6548	Plasmid pTS174 use
9	188.8	45.5	6548	Plasmid pTS172. C
10	188.8	45.5	6548	E. coli plasmid pT
11	188.8	45.5	7492	Plasmid pTS346. U

c 12	188.4	45.4	1303	17	AAT39337	Plasmid pTS88 (Eco
c 13	188.4	45.4	3201	12	AAQ14529	pFS029 Bt ICP codi
c 14	188.4	45.4	4832	22	AAH25423	Nucleotide sequenc
c 15	188.4	45.4	4946	18	AAT59531	T-DNA of plasmid p
c 16	188.4	45.4	4946	22	AAH25422	Nucleotide sequenc
c 17	188.4	45.4	5349	19	AAV23239	T-DNA of pTS24.
c 18	188.4	45.4	5864	17	AAT39339	Plasmid pCOLI3 T-
c 19	188.4	45.4	5865	22	AAH25439	Chimeric T-DNA of
c 20	188.4	45.4	7566	14	AAQ42160	Plasmid pFS0212 co
c 21	188.4	45.4	7639	14	AAQ42159	Plasmid pJD884 con
c 22	182.4	44.0	1037	11	AAQ04705	USP-Promoter-casse
c 23	182.4	44.0	1085	11	AAQ04703	Legumin-signalpept
c 24	182.4	44.0	1160	11	AAQ04706	USP-signalpeptide
c 25	179.4	43.2	1077	22	AAH25439	Right flanking reg
c 26	177	42.7	3201	12	AAQ15144	pVE36 Bt ICP codin
c 27	153	36.9	1186	13	AAQ25707	Chimeric neo gene
c 28	146	35.2	3153	21	AAZ29122	Plasmid DV131 comp
c 29	146	35.2	3336	21	AAZ29121	Plasmid DV130 comp
c 30	146	35.2	3694	21	AAZ29124	Plasmid DV133 used
c 31	146	35.2	3877	21	AAZ29123	Plasmid DV132 used
c 32	146	35.2	24593	6	AAH50226	Sequence of opine
c 33	146	35.2	24596	6	AAH50182	Complete nucleotid
c 34	107.6	25.9	936	22	AAF58252	Oligonucleotide D1
c 35	107.6	25.9	936	22	AAF58254	Oligonucleotide D1
c 36	107.6	25.9	936	22	AAF58257	Oligonucleotide D1
c 37	107.6	25.9	936	22	AAF58259	Oligonucleotide D2
c 38	107.6	25.9	936	22	AAF58262	Oligonucleotide D2
c 39	107.6	25.9	938	22	AAF58255	Oligonucleotide D1
c 40	106.4	25.6	936	22	AAF58252	Oligonucleotide D1
c 41	106.4	25.6	936	22	AAF58254	Oligonucleotide D1
c 42	106.4	25.6	936	22	AAF58257	Oligonucleotide D1
c 43	106.4	25.6	936	22	AAF58259	Oligonucleotide D2
c 44	106.4	25.6	936	22	AAF58262	Oligonucleotide D2
c 45	106.4	25.6	938	22	AAF58255	Oligonucleotide D1

ALIGNMENTS

RESULT 1  
AAD06997  
ID AAD06997 standard; DNA; 415 BP.  
XX  
AC AAD06997;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE Right (5') border flanking region of elite event MS-B2.  
XX  
KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; ds.  
XX  
OS Chimeric - Agrobacterium sp.  
OS Chimeric - Brassica sp.  
FH Key Location/Qualifiers  
FT misc\_feature 1..234  
FT /tag= a  
FT /note= "Corresponds to plant DNA"  
FT misc\_feature 235..415  
FT /tag= b  
FT /note= "Corresponds to T-DNA"  
XX  
WO200131042-A2.  
PN  
PD 03-MAY-2001.  
XX  
PF 26-OCT-2000; 2000WO-EP10680.  
XX  
PR 29-OCT-1999; 99US-0430497.  
XX  
(AVET ) AVENTIS CROPS SCIENCE NV.  
XX





```

RESULT 3
AAT39339/c
ID AAT39339 standard; DNA; 5864 BP.
XX AC AAT39339;
XX DT 22-JAN-1997 (first entry)
XX XX Plasmid pTCOL13 T-DNA used to obtain male sterile oilseed rape.
DE DE Plasmid pTCOL13; male sterile; barnase; ribonuclease; barstar;
KW KW transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX XX Synthetic.
XX OS
XX FH Key Location/Qualifiers
XX FT misc_feature
XX FT complement (1..25)
XX FT /*tag= a
XX FT /label= RB
XX FT /note= "right border of Agrobacterium T-DNA"
XX FT polyA_signal
XX FT complement (98..330)
XX FT /*tag= b
XX FT /label= 3'g7
XX FT /note= "region containing polyA signal of gene 7
XX FT of Agrobacterium T-DNA"
XX FT CDS
XX FT complement (331..882)
XX FT /*tag= c
XX FT /label= bar
XX FT /note= "region coding for phosphinothricin
XX FT acetyltransferase"
XX FT promoter
XX FT complement (883..2608)
XX FT /*tag= d
XX FT /label= Pssu
XX FT /note= "promoter of Arabidopsis Rubisco small
XX FT subunit gene"
XX FT polyA_signal
XX FT complement (2659..3031)
XX FT /*tag= e
XX FT /label= 3'nos
XX FT /note= "region containing polyA signal of nopaline
XX FT synthase gene of Agrobacterium T-DNA"
XX FT CDS
XX FT complement (3032..3367)
XX FT /*tag= f
XX FT /label= Barnase
XX FT /note= "Bacillus amyloliquefaciens barnase coding
XX FT region"
XX FT promoter
XX FT complement (3368..4877)
XX FT /*tag= g
XX FT /label= PTA29
XX FT /note= "promoter of stamen-specific TA29 gene of
XX FT Nicotiana tabacum"
XX FT promoter
XX FT 4924..5216
XX FT /*tag= h
XX FT /label= Pnos
XX FT /note= "promoter of nopaline synthase gene of
XX FT Agrobacterium T-DNA"
XX FT CDS
XX FT 5217..5489
XX FT /*tag= i
XX FT /label= Barstar
XX FT /note= "region coding for barstar of Bacillus
XX FT amyloliquefaciens"
XX FT polyA_signal
XX FT 5490..5765
XX FT /*tag= j
XX FT /label= 3'g7
XX FT /note= "region containing polyA signal of gene 7
XX FT of Agrobacterium T-DNA"
XX FT misc_feature
XX FT complement (5840..5864)
XX FT /*tag= k
XX FT /label= LB
XX FT /note= "left border of Agrobacterium T-DNA"
XX PN W09626283-A1.
XX XX 29-AUG-1996.

```

```

XX 21-FEB-1996; 96WO-EP00722.
XX 21-FEB-1995; 95EP-0400364.
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX Botterman J, Cornelissen M, Michiels F;
XX WPI; 1996-402373/40.
XX Prodn. of male sterile plants by transforming with a chimaeric
XX construct - comprising a male sterility DNA e.g. barnase and a
XX co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX for generating hybrid cultivars
XX Example 3; Page 33-3743-47; 56pp; English.
XX Plasmid pTCOL13 (AAT39339) is a T-DNA vector containing a bar gene
XX under control of the PSU promoter, a barnase gene under control
XX of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
XX gene under control of the Pnos promoter. 87% Of oilseed rape
XX plants regenerated after Agrobacterium-mediated transformation
XX using pTCOL13 were male sterile. Barnase expression disturbed the
XX function of stamen cells leading to male sterility. Constitutive
XX expression of barstar counteracted any low level expression of
XX barnase in non-stamen tissue.
XX Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;
XX
XX Query Match 46.1%; Score 191.4; DB 17; Length 5864;
XX Best Local Similarity 99.5%; Pred. No. 1.4e-30;
XX Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 223 gatgtacatggccgataagaagcaattgttagatgttaattcccatcttgaagaaa 282
Db 5812 GATGTACATGGTCGATATAAGAAAGGCAATTTGTAGATGTTAATTCCTTGAAGAAA 5753
Qy 283 tatagtttaaatattattgataaaataacaagtcagattattatagccaagcaaaa 342
Db 5752 TATAGTTTAAATATTATTGATAAAATAACAAGTCAGGATTTATTATAGTCCAA 5693
Qy 343 taaattattgatgcaagtttaaatcagaataattccaataactgattatcagctgg 402
Db 5692 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGG 5633
Qy 403 tacattgccgtag 415
Db 5632 TACATTGCCGTAG 5620
XX
RESULT 4
AAD06990/c
ID AAD06990 standard; DNA; 5865 BP.
XX AC AAD06990;
XX DT 06-AUG-2001 (first entry)
XX XX Chimeric T-DNA of plasmid pTCOL13.
XX T-DNA: plasmid pTCOL13; transgenic Brassica plant; transformation event;
XX male-sterility gene; chimeric; tobacco; ds.
XX Chimeric - Streptomyces hygroscopicus.
XX Chimeric - Arabidopsis thaliana.
XX Chimeric - Bacillus amyloliquefaciens.
XX Chimeric - Nicotiana tabacum.
XX Chimeric - Agrobacterium tumefaciens.
XX Chimeric - Unidentified.
XX Key Location/Qualifiers
XX FH

```



```
Oy 403 tacattgccgtag 415
Db 5633 TACATTGCCGTAG 5621

RESULT 5
AAF25320/C
ID AAF25320 standard; DNA; 7599 BP.
XX
AC AAF25320;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plasmid PGKB5.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Synthetic.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR01768.
XX
PR 25-JUN-1999; 99FR-0008185.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Hoffmann B, Mollier P, Pelletier G;
XX
DR WPI; 2001-102893/11.
XX
CC New constitutive plant promoter active specifically in roots, useful
PT for controlling expression of pest or drought resistance genes, and
PT related transgenic plants -
XX
PS Disclosure; Fig 9; 92pp; French.
XX
CC The present sequence represents a plasmid PGKB5. The plasmid contains
CC a plant promoter that directs expression of a selected sequence in
CC root cells at all stages of development of a plant. The plant promoter
CC is used to control expression of genes in a root-specific manner,
CC especially genes that provide resistance to parasites, pests (nematodes
CC or fungi), water and salt stress, or alter sugar content or nitrogen
CC transport. Fragments of the promoter are useful as probes or primers
CC to detect or amplify at least part of the promoter.
XX
SQ Sequence 7599 BP; 1972 A; 1938 C; 1937 G; 1752 T; 0 other;

Query Match 46.18; Score 191.4; DB 22; Length 7599;
Best Local Similarity 99.58; Pred. No. 1.4e-30;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 223 gatgtacatggccgataagaagcaattgttagatgttaattcccatcttgaaagaa 282
Db 6771 GATGTACATGTCGATAGAAGAAAGCAATTTGTAGATGTTAATTCCTTCATGAAAGAAA 6712

Oy 283 tatagtttaaatattttatgataaaataacaagtcaggtattatagtcacaagcaaaaaa 342
Db 6711 TATAGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACA 6652

Oy 343 taaatttattgatgcaggttttaattccagaataatttcaataactgattatcatcagctgg 402
Db 6651 TAAATTTATTGATGCAAGTTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGG 6592

Oy 403 tacattgccgtag 415
Db 6591 TACATTGCCGTAG 6579
```

```
RESULT 6
AAF86439/C
ID AAF86439 standard; DNA; 5228 BP.
XX
AC AAF86439;
XX
DT 25-JUN-2001 (first entry)
XX
DE Plasmid pTSL172delta.
XX
KW Male sterile plant; RNAase inhibitor; plasmid pTSL172delta; ds.
XX
OS Unidentified.
XX
PN WO200124616-A1.
XX
PD 12-APR-2001.
XX
PF 12-SEP-2000; 2000WO-JP06222.
XX
PR 30-SEP-1999; 99JP-0279307.
XX
PA (NISB ) JAPAN TOBACCO INC.
XX
PI Hamada K, Nakakido F;
XX
DR WPI; 2001-266212/27.
XX
CC Method for producing male sterile rice and maize by inserting RNase
PT gene and RNase inhibitor genes with promoters into the plant genome -
PT Disclosure; Page 14-17; 29pp; Japanese.
XX
CC The present invention relates to a method for producing male sterile
CC plants. The method comprises inserting a promoter fragment upstream of an
CC RNase gene and a second promoter, upstream of an RNase inhibitor protein
CC gene and inserting it into the plant genome. The method is useful for
CC producing male sterile tobacco, lettuce and rapeseed plants, but
CC preferably rice and maize. The present sequence is a vector used in
CC the method of the present invention.
XX
SQ Sequence 5228 BP; 1384 A; 1307 C; 1263 G; 1274 T; 0 other;

Query Match 45.58; Score 188.8; DB 22; Length 5228;
Best Local Similarity 99.08; Pred. No. 4.6e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 224 atgtacatggccgataagaagcaattgttagatgttaattcccatcttgaaagaaat 283
Db 5222 ACGTACATGTCGATAAGAAAGCAATTTGTAGATGTTAATTCCTTCATCTTGAAAGAAAT 5163

Oy 284 atagtttaaaattatttattgataaaataacaagtcaggtattatagtcacaagcaaaaaa 343
Db 5162 ATAGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACAT 5103

Oy 344 aaatttattgatgcaggttttaattccagaataatttcaataactgattatcatcagctggt 403
Db 5102 AAATTTATTGATGCAAGTTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGGT 5043

Oy 404 acattgccgtag 415
Db 5042 ACATTGCCGTAG 5031

RESULT 7
AAZ91097/C
ID AAZ91097 standard; DNA; 6539 BP.
XX
AC AAZ91097;
XX
DT 06-JUN-2000 (first entry)
```

```

XX DE E. coli plasmid pTS431 containing mutant barnase gene.
XX KW Male sterile plant; mutant barnase gene; anther-specific expression;
KW KW Low fidelity PCR; primer; plant breeding; ss.
XX OS Synthetic.
XX XX
XX PN WO200008176-A1.
XX PD 17-FEB-2000.
XX PF 03-AUG-1999; 99WO-JP04167.
XX PR 04-AUG-1998; 98JP-0220060.
XX PA (NISR) JAPAN TOBACCO INC.
XX PI Hamada K, Nakakido F;
XX DR WPI; 2000-199581/17.
XX PT Mutate barnase gene for efficient construction of plant transformants,
XX PT particularly male sterile plants free from any undesirable characters
XX PT by specifically expressing the gene alone in anther.
XX PS Example 3; Page 23-27; 30pp; Japanese.
XX CC The invention relates to the generation of male sterile plants by
XX CC the introduction of a mutant barnase gene (AA291095) for expression
XX CC specifically in the anther of a plant. This sequence represents the
XX CC E. coli/Agrobacterium shuttle vector plasmid pTS172 which contains
XX CC the mutated barnase gene (AA291095) under control of the cauliflower
XX CC mosaic virus 35S promoter. The vector also contains a region of the
XX CC Agrobacterium T-DNA gene 7. The vector is used for transmitting the
XX CC barnase gene to plants via an Agrobacterium tumefaciens host cell.
XX CC The transformed plant is used in plant breeding.
XX SQ Sequence 6539 BP; 1755 A; 1578 C; 1519 G; 1687 T; 0 other;

Query Match 45.5%; Score 188.8; DB 21; Length 6539;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 224 atgtacatgccgataagaagaagcaatttggtagtggtaattcccatcttgaagaagaat 283
Db 6533 ACGTACATGGTCGATAAGAAAGGCAATTTGTAGATGTTAATTCCTCATCTGAAAGCAAT 6474
Qy 284 atagtttaaatattatttgataaataaacaagtcaggtattatagtcgaagcaaaacat 343
Db 6473 ATAGTTTAAATATTATTGATAAATAACACAGTCAGGTATTATAGTCCAGCAAAACAT 6414
Qy 344 aaatttattgacaggttttaattcagaataatttcgaataactgattatcagctggt 403
Db 6413 AAATTTATTGATCAAGTTTAAATTCAGAAATATTTCATCAACTGATTATATCAGCTGCT 6354
Qy 404 acattgccgtag 415
Db 6353 ACATTGCCGTAG 6342

RESULT 8
AAT39336/C
ID AAT39336 standard; DNA; 6548 BP.
XX AC AAT39336;
XX AC AAT39336;
XX DT 22-JAN-1997 (first entry)
XX DE Plasmid pTS174 used to obtain male sterile rice.
XX KW Plasmid pTS174; male sterile; barnase; ribonuclease; transgenic plant;

```

```

KW Rice; Oryza sativa; ds; cyclic.
XX Synthetic.
XX OS
XX FH Key Location/Qualifiers
XX FT misc_feature 1..2003
XX FT /*tag= a
XX FT /label= Vector
XX FT /note= "pUC19 derived vector sequences"
XX FT complement (2019..2283)
XX FT /*tag= b
XX FT /label= 3'nos
XX FT /note= "region containing polyadenylation signal
XX FT nopaline synthase gene of Agrobacterium
XX FT T-DNA"
XX FT complement (2284..2624)
XX FT /*tag= c
XX FT /label= Barnase
XX FT /product= Bacillus amyloliquefaciens barnase
XX FT complement (2625..4313)
XX FT /*tag= d
XX FT /label= PEI
XX FT /function= promoter of the stamen-specific E1 gene
XX FT 4336..5710
XX FT /*tag= e
XX FT /label= P35S
XX FT /function= 35S promoter of cauliflower mosaic virus
XX FT 5711..6262
XX FT /*tag= f
XX FT /label= bar
XX FT /product= phosphinothricin acetyltransferase
XX FT 6263..6496
XX FT /*tag= g
XX FT /label= 3'g7
XX FT /function= region containing polyadenylation signal
XX XX
XX PN WO9626283-A1.
XX PD 29-AUG-1996.
XX PF 21-FEB-1996; 96WO-EP00722.
XX PR 21-FEB-1995; 95EP-0400364.
XX XX (PLBZ) PLANT GENETIC SYSTEMS NV.
XX PA Botterman J, Cornelissen M, Michiels F;
XX PI WPI; 1996-402373/40.
XX DR prodn. of male sterile plants by transforming with a chimeric
XX XX construct comprising a male sterility DNA e.g. barnase and a
XX PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX PT for generating hybrid cultivars
XX XX Example 1; Page 33-37; 56pp; English.
XX PS Plasmid pTS174 (AAT39336) contains Bacillus barnase DNA under control
XX CC of the stamen-specific PEI promoter. Embryogenic callus from rice
XX CC cv. Kochihibiki was transformed with pTS174 alone or with pTS88
XX CC (see also AAT39337), a plasmid contg. barstar DNA under control of a
XX CC 35S promoter. With pTS174 alone, 1 male sterile line was recovered
XX CC from 48 electroporation cuvettes. With both plasmids, 7 normal
XX CC male sterile lines were recovered from 40 cuvettes. Barnase
XX CC expression disturbed the function of stamen cells leading to male
XX CC sterility. Constitutive expression of barstar counteracted any low
XX CC level expression of barnase in non-stamen tissue.
XX SQ Sequence 6548 BP; 1757 A; 1578 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 17; Length 6548;

```



```

PF 03-AUG-1999; 99WO-JP04167.
XX
PR 04-AUG-1998; 98JP-0220060.
XX
PA (NISR ) JAPAN TOBACCO INC.
XX
PI Hamada K, Nakakido F;
XX WPI; 2000-195581/17.
XX
PT Mutate barnase gene for efficient construction of plant transformants,
PT particularly male sterile plants free from any undesirable characters
PT by specifically expressing the gene alone in anther -
XX
PS Example 3; Page 19-23; 30pp; Japanese.
XX
CC The invention relates to the generation of male sterile plants by
CC the introduction of a mutant barnase gene (AAZ91095) for expression
CC specifically in the anther of a plant. This sequence represents the
CC E. coli/Agrobacterium shuttle vector plasmid pTS172 which contains
CC the synthetic barnase gene (AAZ91094) under control of the cauliflower
CC mosaic virus 35S promoter. The vector also contains a region of the
CC Agrobacterium T-DNA gene 7. The vector is used for transmitting the
CC barnase gene to plants via an Agrobacterium tumefaciens host cell.
CC The transformed plant is used in plant breeding.
XX
SQ Sequence 6548 BP; 1756 A; 1579 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 21; Length 6548;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatggccgataagaagaaggcaattttagatgttaattcccatcttgaaagaaat 283
DB 6542 ACGTACATGGTTCGATAGAGAAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAAT 6483
QY 284 atagtttaaatatttattgataaaatacaagtcaggattattatagtcacagcaaaacat 343
DB 6482 ATAGTTTAAATATTATTGATTAATAATACCAAGTCAGGTATTATAGTCCCAAGCAAAACAT 6423
QY 344 aaattatttgatgcagtttaattcagaataatttcaataactgattatcatcagctggt 403
DB 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATTAACATGATTATATCAGCTGGT 6363
QY 404 acattgcccgtag 415
DB 6362 ACATTGCCGTAG 6351

RESULT 11
AAF86441/c
ID AAF86441 standard; DNA; 7492 BP.
XX
AC AAF86441;
XX
DT 25-JUN-2001 (first entry)
XX
DE Plasmid pTS346.
XX
KW Male sterile plant; RNAase inhibitor; plasmid pTS346; ds.
XX
OS Unidentified.
XX
PN WO200124616-A1.
XX
PD 12-APR-2001.
XX
PF 12-SEP-2000; 2000WO-JP06222.
XX
PR 30-SEP-1999; 99JP-0279307.
XX
PA (NISR ) JAPAN TOBACCO INC.

```

```

XX Hamada K, Nakakido F;
XX WPI; 2001-266212/27.
XX
PT Method for producing male sterile rice and maize by inserting RNase
PT gene and RNase inhibitor genes with promoters into the plant genome -
XX Disclosure; Page 19-23; 29pp; Japanese.
XX
CC The present invention relates to a method for producing male sterile
CC plants. The method comprises inserting a promoter fragment upstream of an
CC RNase gene and a second promoter, upstream of an RNase inhibitor protein
CC gene and inserting it into the plant genome. The method is useful for
CC producing male sterile tobacco, lettuce and rapeseed plants, but
CC preferably rice and maize. The present sequence is a vector used in
CC the method of the present invention.
XX
SQ Sequence 7492 BP; 1987 A; 1801 C; 1752 G; 1952 T; 0 other;

Query Match 45.5%; Score 188.8; DB 22; Length 7492;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatggccgataagaagaaggcaattttagatgttaattcccatcttgaaagaaat 283
DB 7486 ACGTACATGGTTCGATAGAGAAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAAT 7427
QY 284 atagtttaaatatttattgataaaatacaagtcaggattattatagtcacagcaaaacat 343
DB 7426 ATAGTTTAAATATTATTGATTAATAATACCAAGTCAGGTATTATAGTCCCAAGCAAAACAT 7367
QY 344 aaattatttgatgcagtttaattcagaataatttcaataactgattatcatcagctggt 403
DB 7366 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATTAACATGATTATATCAGCTGGT 7307
QY 404 acattgcccgtag 415
DB 7306 ACATTGCCGTAG 7295

RESULT 12
AAT39337/c
ID AAT39337 standard; DNA; 1303 BP.
XX
AC AAT39337;
XX
DT 22-JAN-1997 (first entry)
XX
DE Plasmid pTS88 (EcoRI-HindIII fragment).
XX
KW Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1..35
FT /*tag= a
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
FT promoter 36..694
FT /*tag= b
FT /label= p35S
FT /function= 35S promoter of cauliflower mosaic virus
FT CDS 695..967 strain CM1841
FT /*tag= c
FT /label= barstar
FT /product= Bacillus amyloliquefaciens barstar
FT polyA_signal 968..1287
FT /*tag= d

```

```

FT      /label= 3'g7
FT      /function= region containing polyadenylation signal
FT      of gene 7 og Agrobacterium T-DNA
FT      misc_feature
FT      1288..1303
FT      /*tag= e
FT      /label= pGEM2
FT      /note= "polylinker of pGEM2"
XX      WO9626283-A1.
XX      29-AUG-1996.
XX      21-FEB-1996; 96WO-EP007722.
XX      21-FEB-1995; 95EP-0400364.
XX      (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX      Botterman J, Cornelissen M, Michiels F;
XX      WPI; 1996-402373/40.
XX      Prodn. of male sterile plants by transforming with a chimaeric
XX      construct - comprising a male sterility DNA e.g. barnase and a
XX      co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX      for generating hybrid cultivars
XX      Example 1; Page 38; 56pp; English.
XX      The HindIII-EcoRI fragment (AA139337) of plasmid pTS88 contains
XX      barstar DNA under control of a 35S promoter. The plasmid was
XX      used with pTS174 (see also AA139336) contg. barnase DNA under
XX      control of the stamen-specific promoter E1 to produce male sterile
XX      rice cv. Kochihibiki transgenic plants, and with plasmid pV136
XX      (see also AA139338) contg. barnase DNA under control of the stamen-
XX      specific PCA55 promoter to produce male sterile maize plants.
XX      Expression of barnase (a ribonuclease) in the stamen leads to male
XX      sterility. Constitutive expression of barstar counteracts possible
XX      low level expression of barnase DNA in non-stamen tissue.
XX      Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;
XX      Query Match 45.4%; Score 188.4; DB 17; Length 1303;
XX      Best Local Similarity 99.5%; Pred. No. 5e-30;
XX      Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      226 gtacatggccgataagaagaagcaatttctagatgttaattcccatcttgaagaataat 285
DB      1287 GTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTTGAAAGAAATAT 1228
QY      286 agtttaaatattattgataaaataacaagtcaggtattatagtcgaagcaaaaacataa 345
DB      1227 AGTTTAAATATTATTGATATAAATAACAGTCAGGTATTATAGTCCCAAGCAAAAACATAA 1168
QY      346 atttattgatgcaagtttaaatcagaataatttcaataactgattatcacgtgggtac 405
DB      1167 ATTTATTGTCGAAGTTTAAATTCAGAAATATTTCATTAACACTGATTATATCAGCTGGTAC 1108
QY      406 attgccgtag 415
DB      1107 ATTGCCGTAG 1098
XX      RESULT 13
XX      AAQ14529/c
ID      AAQ14529 standard; DNA; 3201 BP.
XX      AAQ14529;
XX      27-JAN-1992 (first entry)
XX      pp5029 Bt ICP coding sequence.

```

```

XX      Bacillus thuringiensis; insecticidal crystal protein; ICP;
KW      deletion; ss.
XX      Synthetic.
XX      WO9116432-A.
XX      31-OCT-1991.
XX      17-APR-1991; 91WO-EP007733.
XX      18-APR-1990; 90EP-0401055.
XX      (PLAN-) PLANT GENETIC SYST.
XX      Cornelissen M, Soetaert P, Stam M, Dockx J;
XX      WPI; 1991-339820/46.
XX      Modified Bacillus thuringiensis insecticidal crystal protein
XX      genes - having A and T sequences changed to G and C sequences
XX      encoding same amino acids, for increased expression levels
XX      Disclosure; Fig 6(c); 78pp; English.
XX      "n" in the sequence refers to not known nucleotides.
XX      pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-
XX      terminal modification and the internal modification of the Bt ICP
XX      coding sequence.
XX      See also AAQ14529, AAQ15142-44.
XX      Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;
XX      Query Match 45.4%; Score 188.4; DB 12; Length 3201;
XX      Best Local Similarity 99.5%; Pred. No. 5.3e-30;
XX      Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      226 gtacatggccgataagaagaagcaatttctagatgttaattcccatcttgaagaataat 285
DB      3141 GTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTTGAAAGAAATAT 3082
QY      286 agtttaaatattattgataaaataacaagtcaggtattatagtcgaagcaaaaacataa 345
DB      3081 AGTTTAAATATTATTGATATAAATAACAGTCAGGTATTATAGTCCCAAGCAAAAACATAA 3022
QY      346 atttattgatgcaagtttaaatcagaataatttcaataactgattatcacgtgggtac 405
DB      3021 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATTAACACTGATTATATCAGCTGGTAC 2962
QY      406 attgccgtag 415
DB      2961 ATTGCCGTAG 2952
XX      RESULT 14
XX      AAH25423
ID      AAH25423 standard; DNA; 4832 BP.
XX      AAH25423;
XX      22-AUG-2001 (first entry)
XX      Nucleotide sequence of plasmid pTHW118.
XX      Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW      fertility restorer gene; barstar gene; ss.
XX      Synthetic.
XX      Streptomyces hygroscopicus.
XX      Arabidopsis thaliana.
XX      Bacillus amyloliquefaciens.

```

```

OS Nicotiana tabacum.
XX
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "right border repeat from TL-DNA from pTIB6S3"
FT 26..53
FT /tag= b
FT /note= "synthetic polylinker derived sequences"
FT 54..90
FT /tag= c
FT /note= "residual sequence from TL-DNA at right border repeat"
FT 91..97
FT /tag= d
FT /note= "synthetic polylinker derived sequences"
FT 98..309
FT /tag= e
FT /note= "3' UTR from TL-DNA gene 7 of pTIB6S3"
FT 310..330
FT /tag= f
FT /note= "synthetic polylinker derived sequences"
FT complement (331..882)
FT /tag= g
FT /note= "Streptomyces hygroscopicus bialaphos resistance (bar) gene"
FT complement (883..2608)
FT /tag= h
FT /note= "atsLA ribulose-1,5-biphosphate carboxylase small subunit gene from Arabidopsis thaliana"
FT 2609..2658
FT /tag= i
FT /note= "synthetic polylinker derived sequences"
FT complement (2659..2919)
FT /tag= j
FT /note= "Tag1 fragment from 3' UTR of nopaline synthase gene from T-DNA of pTiT37 and containing plant polyadenylation signals"
FT 2920..2940
FT /tag= k
FT /note= "synthetic polylinker derived sequences"
FT 2941..2980
FT /tag= l
FT /note= "downstream of Bacillus amyloliquefaciens barstar coding region"
FT complement (2981..3253)
FT /tag= m
FT /note= "Barstar gene coding region from Bacillus amyloliquefaciens"
FT complement (3254..4762)
FT /tag= n
FT /note= "anther-specific gene TA29 promoter from Nicotiana tabacum"
FT 4763..4807
FT /tag= o
FT /note= "synthetic polylinker derived sequences"
FT 4808..4832
FT /tag= p
FT /note= "left border repeat from TL-DNA from pTIB6S3"
FT
XX
PN WO200141558-A1.
XX
XX 14-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-EP12872.
XX
XX 08-DEC-1999; 99US-0457037.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX De Both G, De Beuckeleer M;
XX
XX WPI; 2001-381419/40.

```

```

XX Transgenic winter oilseed rape plants suited for producing hybrid seed
PT with improved qualities, comprises a male-sterility gene and fertility
PT restorer gene, integrated into the genome -
XX
XX Example 1; Page 80-82; 98pp; English.
XX
CC The specification describes a pair of transgenic winter oilseed rape
CC plants suited for producing hybrid seed. One of the plants has an
CC expression cassette comprising a male-sterility gene, and the other
CC plant has an expression cassette comprising a fertility restorer gene,
CC integrated into the genome. The fertility restorer gene is capable of
CC preventing the activity of the male-sterility gene. The plant pair is
CC useful for producing hybrid seed. Plants developed from the hybrid
CC seed have agronomic performance, genetic stability and adaptability to
CC different genetic backgrounds. The present sequence represents
CC a plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
CC a fertility restorer gene. The plasmid is used to create transgenic
CC plants of the invention.
XX
SQ Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;

Query Match 45.4%; Score 188.4; DB 22; Length 4832;
Best Local Similarity 99.5%; Pred. No. 5.5e-30;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatggccgatagaagaagcaattttagatgttaattcccatcttgaagaataat 285
Db 54 gtacatggcgcgaagaagaagcaattttagatgttaattcccatcttgaagaataat 113
QY 286 agtttaaatattattgataaaataacaaagtcagggtattatagtcacagcaaaacataa 345
Db 114 agtttaaatattattgataaaataacaaagtcagggtattatagtcacagcaaaacataa 173
QY 346 atttattgacgaagttaaatcagaataatttcaataactgattatcagctggtac 405
Db 174 atttattgacgaagttaaatcagaataatttcaataactgattatcagctggtac 233
QY 406 attgcgcgtag 415
Db 234 attgcgcgtag 243

RESULT 15
AAT59531
ID AAT59531 standard; DNA; 4946 BP.
AC AAT59531;
XX
XX 07-MAY-1997 (first entry)
XX
XX T-DNA of plasmid pTHW107.
XX
XX Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PARP;
KW nicinamide; Agrobacterium; T-DNA; male sterile; barnase;
KW ribonuclease; RNase; cereal; wheat; oilseed rape; Brassica napus;
KW plasmid pTHW107; ds.
XX
XX Chimeric Agrobacterium sp.;
OS Chimeric Arabidopsis thaliana;
OS Chimeric Nicotiana tabacum.
XX
XX Key Location/Qualifiers
FT misc_RNA complement (1..25)
FT /tag= a
FT /label= RB
FT /note= "T-DNA right border"
FT 3'UTR complement (97..330)
FT /tag= b
FT /label= 3'g7
FT /note= "3' untranslated region contg. the poly-A signal of gene-7 of Agrobacterium T-DNA"
FT

```





\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:16 ; Search time 301.6 seconds  
(without alignments)  
311.633 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagtggtgttcattga.....cagctgtacattgcgtag 415

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	191.4	46.1	5864	3	US-08-894-440-4
C 2	188.8	45.5	6548	3	US-08-894-440-1
C 3	188.8	45.5	6548	3	US-08-817-188-2
C 4	188.8	45.5	7811	2	US-08-549-680A-5
C 5	188.4	45.4	1303	3	US-08-894-440-2
C 6	188.4	45.4	3200	1	US-08-453-104-23
C 7	188.4	45.4	3200	2	US-08-694-824-23
C 8	188.4	45.4	4946	3	US-08-817-188-1
C 9	188.4	45.4	5560	3	US-08-817-188-5
C 10	188.4	45.4	7566	2	US-08-232-016-23
C 11	188.4	45.4	7639	2	US-08-232-016-22
C 12	186.8	45.0	5864	3	US-08-894-440-4
C 13	177	42.7	3201	1	US-08-453-104-22
C 14	177	42.7	3201	2	US-08-694-824-22
C 15	153	36.9	1186	1	US-08-064-121-2
C 16	153	36.9	1186	1	US-08-478-015-2
C 17	153	36.9	1186	3	US-08-475-975-2
C 18	153	36.9	1186	3	US-09-084-889-2
C 19	146	35.2	3153	4	US-09-080-625-3
C 20	146	35.2	3336	4	US-09-080-625-2
C 21	146	35.2	3694	4	US-09-080-625-5
C 22	146	35.2	3877	4	US-09-080-625-4
C 23	143	34.5	24595	6	5428147-1
C 24	44.8	10.8	8654	1	US-08-920-812-6
C 25	44.8	10.8	8654	1	US-08-920-827-6
C 26	44.8	10.8	8654	1	US-08-921-177-6
C 27	44.8	10.8	8654	1	US-08-362-577C-6

28	44.8	10.8	8654	2	US-08-920-828-6	Sequence 6, Appli
C 29	44.6	10.7	5526	3	US-08-751-359-21	Sequence 21, Appl
C 30	43.6	10.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 31	43	10.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 32	42.6	10.3	8920	2	US-08-446-855A-1	Sequence 1, Appli
C 33	42.6	10.3	8920	4	US-09-150-741-1	Sequence 1, Appli
C 34	42	10.1	1316	2	US-08-871-924A-1	Sequence 1, Appli
C 35	40.4	9.7	1588	3	US-09-058-489-45	Sequence 45, Appli
C 36	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appli
C 37	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appli
C 38	39.8	9.6	24595	6	5428147-1	Patent No. 5428147
C 39	39.4	9.5	80595	4	US-09-078-294-3	Sequence 3, Appli
C 40	39.2	9.4	658	4	US-08-998-416-595	Sequence 595, App
C 41	38.8	9.3	636	4	US-08-998-416-1137	Sequence 1137, Ap
C 42	38.8	9.3	660	1	US-07-991-867B-32	Sequence 32, Appl
C 43	38.8	9.3	660	1	US-08-107-755A-32	Sequence 32, Appl
C 44	38.8	9.3	660	2	US-08-544-332-32	Sequence 32, Appl
C 45	38.8	9.3	1511	1	US-07-991-867B-8	Sequence 8, Appli

ALIGNMENTS

```
RESULT 1
US-08-894-440-4/c
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894.440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTCO113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: napalase synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: anlyoliquefaciens
; FEATURE:
```

```

; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
;
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
;
US-08-894-440-4

Query Match          46.1%; Score 191.4; DB 3; Length 5864;
Best Local Similarity 99.5%; Pred.No. 2.7e-37;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps

Qy 223 gatgtacatgcccgcataagaagaaaggccaattgtagatgttaattcccatcttgaagaaa 282
Db 5812 GATGTACATGCTCGATAAGAAAAGGCAATTTCGTAGATCTTAATCCCATCTTGAAAAGAAA 5753
Qy 283 tataagtttaaatatttatgataaatacaaaagtcagggtatattagttccaaagcaaaaaca 342
Db 5752 TAGATGTTAAATATTATTGATATAAATACAAAGTCAGGTATTATAGTCCAAAGCAAAAACA 5693
Qy 343 taaattattgatgcgaagtttaaatccagaaatattcaataactgattatatacagctgg 402
Db 5692 TAAATTATTGATGCAAGTTTAATTCAGAAATATTTCAATACTGATTATATCAGCTGG 5633
Qy 403 tacatgcccgcag 415
Db 5632 TACATTGCCGTAG 5620

RESULT 2
US-08-894-440-1/c
; Sequence 1, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTSL174
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2003)
; OTHER INFORMATION: pUC19 derived vector sequences (vector)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2283))
; OTHER INFORMATION: 3' nos: region containing polyadenylation signal

```

```

; LOCATION: Complement((2019)..(2288))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2289)..(2624))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: PE1: promoter region of E1 gene of rice
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5170)
; OTHER INFORMATION: P35S: 35S promoter region of Cauliflower Mosaic
; OTHER INFORMATION: Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5711)..(6262))
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; US-08-817-188-2

```

Query Match	45.5%	Score 188.8;	DB 3;	Length 6548;
Best Local Similarity	99.0%	Pred. No. 1.1e-36;		
Matches 190;	Conservative	0;	Mismatches 2;	Indels 0; Gaps 0;
QY 224	atgcacatggccgataagaaaaggcaatttgtagatgttaattcccatctttgaaagaaat	283		
Db 6542	ACGTACATGTCGATAGAAAAGGCANTTGTAGATGTTAATTCCTTGAAGAAGAAAT	6483		
QY 284	atagtttcaatatattatgtatataaaatacaacagtcaggtattatagtcacgaacaaaaacat	343		
Db 6482	ATAGTTTAAATATTTATTGTAAAAATAACAAGTCAGGTATTATTAGTCCAAGCAAAAACAT	6423		
QY 344	aaattttatgtgcgaagtctaaattcagaataatttcaataaactgattatcatcagctgggt	403		
Db 6422	AAATTATTTGATGCAAGGTTAAATTTCAGAAATATTTCAATAACATGATTTATATCAGCTGGT	6363		
QY 404	acattggccgtag	415		
Db 6362	ACATTGCCGTAG	6351		

[illegible]

```
US-08-894-440-2/c
; Sequence 2, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
; OTHER INFORMATION: fragment of pTS88
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(35)
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36)..(694)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain
; OTHER INFORMATION: CM1841 (P35S)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (695)..(967)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefians
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (968)..(1287)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1288)..(1303)
; OTHER INFORMATION: polylinker of pGEM2
US-08-894-440-2

Query Match          45.4%; Score 188.4; DB 3; Length 1303;
Best Local Similarity 99.5%; Pred. No. 1.1e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgcataaagaagcaatttctagatgttaattcccatcttgaaagaatat 285
Db 1287 GTACATGTCGTAAGAAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 1228

QY 286 agtttaaatatttattgataaaaaaacagtcagggtattatagtcacagcaaaacataa 345
Db 1227 AGTTTAAATATTATTGATAAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACATAA 1168

QY 346 atttattgatgcaggttttaattcgaataatttcaataactgattatcagctggtac 405
Db 1167 ATTTATTGATGCAAGTTTAAATTTCAGAAATATTTCATAACTGATTATATCAGCTGGTAC 1108

QY 406 attgcccgtag 415
Db 1107 ATTGCCGTAG 1098

RESULT 6
US-08-453-104-23/c
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETABERT, Piet
; APPLICANT: STAM, Maïke

US-08-894-440-2/c
; Sequence 2, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
; OTHER INFORMATION: fragment of pTS88
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(35)
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36)..(694)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain
; OTHER INFORMATION: CM1841 (P35S)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (695)..(967)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefians
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (968)..(1287)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1288)..(1303)
; OTHER INFORMATION: polylinker of pGEM2
US-08-894-440-2

Query Match          45.4%; Score 188.4; DB 1; Length 3200;
Best Local Similarity 99.5%; Pred. No. 1.3e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgcataaagaagcaatttctagatgttaattcccatcttgaaagaatat 285
Db 3141 GTACATGTCGTAAGAAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 3082

QY 286 agtttaaatatttattgataaaaaaacagtcagggtattatagtcacagcaaaacataa 345
Db 3081 AGTTTAAATATTATTGATAAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACATAA 3022

QY 346 atttattgatgcaggttttaattcgaataatttcaataactgattatcagctggtac 405
Db 3021 ATTTATTGATGCAAGTTTAAATTTCAGAAATATTTCATAACTGATTATATCAGCTGGTAC 2962

QY 406 attgcccgtag 415
Db 2961 ATTGCCGTAG 2952

RESULT 7
US-08-694-824-23/c
; Sequence 23, Application US/08694824
; Patent No. 5877306
```

```

; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,824
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
US-08-694-824-23

```

```

Query Match 45.4%; Score 188.4; DB 2; Length 3200;
Best Local Similarity 99.5%; Pred. No. 1.3e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatgccgataagaagcaattgttagatgtaattcccatcttgaaagaataat 285
Db 3141 GTACATGTCGATAGAAAAGCGCAATTGTGTAGATGTTAAATCCCATCTTGAAGAAATAT 3082

Qy 286 agttaaataattattgataaaaataacaagtcaggattattagtcacagcaaaaacataa 345
Db 3081 AGTTTAAATATTATTGATATAAAATAACAAGTCAGGTATTATAGTCCAGCAAAAACATAA 3022

Qy 346 atttattgatgcaggttttaattcagaataatttcaataactgattatcatcagctggtac 405
Db 3021 ATTATTATGATCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGTAC 2962

Qy 406 attgccgtag 415
Db 2961 ATTGCCGTAG 2952

```

```

RESULT 8
US-08-817-188-1
; Sequence 1, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4946
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW107
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: T-DNA right border (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((97)..(330))
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl
; OTHER INFORMATION: transferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter region of Rubisco small subunit gene of
; OTHER INFORMATION: Arabidopsis thaliana (PSSU)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2658)..(3031))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4876))
; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((4922)..(4946))
; OTHER INFORMATION: LB: T-DNA left border
US-08-817-188-1

```

```

Query Match 45.4%; Score 188.4; DB 3; Length 4946;
Best Local Similarity 99.5%; Pred. No. 1.4e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatgccgataagaagcaattgttagatgtaattcccatcttgaaagaataat 285
Db 54 gtacatggtcgataagaagcaattgttagatgtaattcccatcttgaaagaataat 113

```

```

QY 286 agtttaaatattattgataaaataacaagtcaggtattattagtcacaaagcaaaacataa 345
Db 114 agtttaaatattattgataaaataacaagtcaggtattattagtcacaaagcaaaacataa 173
QY 346 atttattgatcaagtttaaatcagaataattttcaaatcatttatcagctgggtac 405
Db 174 atttattgatcaagtttaaatcagaataattttcaaatcatttatcagctgggtac 233
QY 406 attgccgtag 415
Db 234 attgccgtag 243

RESULT 9
US-08-817-188-5
; Sequence 5, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' q7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-Ls1 gene
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (5067)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; OTHER INFORMATION: nucleotides)
US-08-817-188-5

Query Match 45.4%; Score 188.4; DB 3; Length 5560;
Best Local Similarity 99.5%; Pred. No. 1.4e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatggccgataaagaaagcaattttagatgtgtaattcccatcttgaagaataat 285
Db 41 gtacatggctcgataaagaaagcaattttagatgtgtaattcccatcttgaagaataat 100
QY 286 agtttaaatattattgataaaataacaagtcaggtattattagtcacaaagcaaaacataa 345
Db 101 agtttaaatattattgataaaataacaagtcaggtattattagtcacaaagcaaaacataa 160
QY 346 atttattgatcaagtttaaatcagaataattttcaaatcatttatcagctgggtac 405
Db 161 atttattgatcaagtttaaatcagaataattttcaaatcatttatcagctgggtac 220
QY 406 attgccgtag 415
Db 221 attgccgtag 230

RESULT 10
US-08-232-016-23/c
; Sequence 23, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; TITLE OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016

```



```
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91402920.2
; FILING DATE: 30-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pPS0212
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated modified bt2 (cryIab) gene, also designated as the
; OTHER INFORMATION: cryIab6 gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1793..2026
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2396..2921
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2922..3581
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3582..4407
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5600..6457
; OTHER INFORMATION: /note= "Sequence complementary to
; OTHER INFORMATION: Patent No. 5952547
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7071..7566
; OTHER INFORMATION: /note= "TR1' and TR2 promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA (with modified leader with
; OTHER INFORMATION: respect to sequence of pJD884 of SEQ ID NO. 22."
; US-08-232-016-23
;
; Query Match 45.4%; Score 188.4; DB 2; Length 7566;
; Best Local Similarity 99.5%; Pred. No. 1.5e-36;
; Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 226 gtacatggccgataagaaggaatttgtagattgaattcccatcttgaagaataat 285
; ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 2070 GTACATGGTCGATAGAGAAAAGCAATTTCTAGATGTTAATTCCCATCTTGAAGAAATAT 2011
Qy 286 agtttaaatatttattgataaaatacaagtcaggtattattatgaccaagcaaaaacataa 345
; ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db 2010 ACTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACATAA 1951
Qy 346 atttattgatgcaagtttaatttaattcagaataatttcaataaactgattatatcagctggtac 405
; ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1950 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTTCAATAACTGATTATATATCAGCTGGTAC 1891
Qy 406 attgccgtag 415
; ||||| |||||||
Db 1890 ATTGCCGTAG 1881
;
; RESULT 11
; US-08-232-016-22/c
; Sequence 22, Application US/082332016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; TITLE OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91402920.2
; FILING DATE: 30-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pJD884
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1869
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated bt2 (cryIab) gene, also designated as the bt884
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: 1877..2110
; OTHER INFORMATION: /note= '3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2480..3005
; OTHER INFORMATION: /note= '35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3006..3665
; OTHER INFORMATION: /note= 'Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3666..4491
; OTHER INFORMATION: /note= '3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5684..6541
; OTHER INFORMATION: /note= 'Sequence complementary to
; Patent No. 5952547
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7155..7639
; OTHER INFORMATION: /note= 'TR1' and TR2' promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA."
US-08-232-016-22
```

```
Query Match 45.4%; Score 188.4; DB 2; Length 7639;
Best Local Similarity 99.5%; Pred. No. 1.5e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccataagaagaaggcaattgttagatgtaattcccatcttgaagaataat 285
Db 2154 GTACATGTCGATGAGAAAGGCAATTGTAGATGTTAATTCCTTGAAGAATAAT 2095

QY 286 agtttaaatattattgataaaatacaagtcagggtattatagtcacgaacaaacataa 345
Db 2094 AGTTTAAATATTATTGATATAAAATAACAGTCAGGTATTATAGTCCCAAGCAAAACATAA 2035

QY 346 atttatgatgcaagtttaattcagaataatttcataaactgattatcagctggtac 405
Db 2034 ATTATTGATGCAAGTTTAAATTACAGAAATATTTCATAACTGATTATATCATCGCTGTAC 1975

QY 406 attgcccgtag 415
Db 1974 ATTGCCGTAG 1965
```

```
RESULT 12
US-08-894-440-4
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
```

```
; OTHER INFORMATION: plasmid pTCO113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement(11)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4
```

```
Query Match 45.0%; Score 186.8; DB 3; Length 5864;
Best Local Similarity 98.9%; Pred. No. 3.4e-36;
Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 226 gtacatgcccataagaagaaggcaattgttagatgtaattcccatcttgaagaataat 285
Db 54 gtacatggtcgataagaagaaggcaattgttagatgtaattcccatcttgaagaataat 113

QY 286 agtttaaatattattgataaaatacaagtcagggtattatagtcacgaacaaacataa 345
Db 114 agtttaaatattattgataaaatacaagtcagggtattatagtcacgaacaaacataa 173

QY 346 atttatgatgcaagtttaattcagaataatttcataaactgattatcagctggtac 405
Db 174 atttatgatgcaagtttaattcagaataatttcataaactgattatcagctggtac 233
```

Qy 406 attgccgtag 415  
|||||  
Db 234 attgccgtag 243

RESULT 13  
US-08-453-104-22/c  
; Sequence 22, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 42.7%; Score 177; DB 1; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 239 aagaaaaggcaattgttagatggttaattcccatcttgaaagaaataatagtttaaatattt 298  
|||||  
Db 3201 AAGAAAGGCAATTGTAGATGTTAAATCCCATCTTGAAGAAATATAGTTTAATATT 3142  
Qy 299 attgataaaatacaagtcagggtattatagtcgaagcaaaacataaattattatgca 358  
|||||  
Db 3141 ATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAACATAAAATTTATTGATGCA 3082

Qy 359 agtttaaatcagaataatttcaataactgattatatacagctggtacattgccgtag 415  
|||||  
Db 3081 AGTTTAATTCAGAAATATTTCAATACTAGTATTATATACAGCTGGTACATCCCGTAG 3025

RESULT 14  
US-08-694-824-22/c  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-22

Query Match 42.7%; Score 177; DB 2; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 239 aagaaaaggcaattgttagatggttaattcccatcttgaaagaaataatagtttaaatattt 298  
|||||  
Db 3201 AAGAAAGGCAATTGTAGATGTTAAATCCCATCTTGAAGAAATATAGTTTAATATT 3142  
Qy 299 attgataaaatacaagtcagggtattatagtcgaagcaaaacataaattattatgca 358  
|||||  
Db 3141 ATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAACATAAAATTTATTGATGCA 3082





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:04 ; Search time 8261.74 Seconds  
(without alignments)  
539.777 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 ggcagttgtgttcacga.....cagctgtacattgcgtag 415

Scoring table: IDENTIFY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372899281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estro:\*

7: em\_estrov:\*

8: em\_hic:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_hic:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	63	15.2	734 13 CNS010MP
2	61	14.7	1101 13 CNS0042W
3	57.2	13.8	844 13 CNS03D01
4	57.2	13.8	1101 13 CNS00088
5	57.2	13.8	1101 13 CNS00238
6	55.4	13.3	1101 13 CNS016LI
7	55	13.3	537 13 AQ506817
8	54.8	13.2	987 13 CNS014PQ
9	54.8	13.2	1101 13 CNS003BD
10	54.6	13.2	980 13 CNS000JG1
11	54.6	13.2	1101 13 CNS000EVL
12	54.4	13.1	893 13 CNS013XE

13	54	13.0	905	13	CNS00KHX	AL077798 Drosophil
14	53.4	12.9	500	10	AU087444	AU087444
15	53.4	12.9	970	13	CNS0182A	AL108796 Drosophil
16	53.2	12.8	1001	13	CNS0155H	AL105023 Drosophil
17	53.2	12.8	1101	13	CNS0106X	AL098595 Drosophil
18	53	12.8	1101	13	CNS016LI	AL106896 Drosophil
19	52.6	12.7	734	13	CNS010MP	AL099163 Drosophil
20	52.6	12.7	1013	13	CNS00J7I	AL075824 Drosophil
21	52.6	12.7	1101	13	CNS00EPO	AL069493 Drosophil
22	52.4	12.6	524	13	CNS01U90	AL167541 Tetradon
23	52.4	12.6	996	13	CNS00FUH	AL071063 Drosophil
24	52.2	12.6	639	13	CNS038CX	AL232458 Tetradon
25	51.8	12.5	928	13	CNS00DKY	AL071865 Drosophil
26	51.8	12.5	1101	13	CNS0039G	AL063921 Drosophil
27	51.8	12.5	1101	13	CNS000FC	AL070972 Drosophil
28	51.6	12.4	992	13	CNS0562R	AL322812 Tetradon
29	51.6	12.4	1043	13	CNS0145P	AL103735 Drosophil
30	51.6	12.4	1101	13	CNS00EO7	AL069440 Drosophil
31	51.4	12.4	876	13	CNS009G1	AL053529 Drosophil
32	51.2	12.3	678	13	CNS02A0C	AL187941 Tetradon
33	51.2	12.3	1001	13	CNS01400	AL103554 Drosophil
34	51.2	12.3	1101	13	CNS00EVL	AL069706 Drosophil
35	50.6	12.2	854	11	BF274512	BF274512 GA_EB002
36	50.6	12.2	963	10	AL566565	AL566565
37	50.6	12.2	1101	13	CNS00CVH	AL060100 Drosophil
38	50.6	12.2	1101	13	CNS0022U	AL097152 Drosophil
39	50.4	12.1	728	13	AQ272964	AQ272964 nbxb0028P
40	50.4	12.1	1101	13	CNS003BB	AL064089 Drosophil
41	50.4	12.1	1200	13	CNS016CO	AL108578 Drosophil
42	50.2	12.1	681	13	B60190	B60190 CIT-HSP-200
43	50	12.0	563	13	AQ326762	AQ326762 nbxb0038D
44	50	12.0	788	13	BH126604	BH126604 BARC-Satt
45	50	12.0	987	13	CNS014PQ	AL104456 Drosophil

ALIGNMENTS

RESULT 1

CNS010MP 734 bp DNA GSS 26-JUL-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC

DEFINITION BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL099163

VERSION AL099163.1 GI:5610774

KEYWORDS GSS,

SOURCE fruit fly.

ORGANISM Plasmid Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 734)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES

source 1.734

Location/Qualifiers

1.734

/organism="Drosophila melanogaster"

/plasmid="pBelOBAC11"

/db\_xref="taxon:7227"

/clone\_lib="DrosBAC"

/clone="BACN04L20"

BASE COUNT	288 a	62 c	2 g	211 t	171 others
ORIGIN	/note="end : 17"				
Query Match	15.2% ; Score 63 ; DB 13 ; Length 734 ;				
Best Local Similarity	41.1% ; Pred. No. 0.0079 ;				
Matches	85 ; Conservative 43 ; Mismatches 79 ; Indels 0 ; Gaps 0 ;				
QY	189	aagttattataataataataataataattatggttaggtatgacatgcccgaataaagcgc 248			
DB	53	AMATATTATATTTATATATAAATATATAAATAAAMMTTCAAAATTTMMAAAAAWAAAW 112			
QY	249	aattgttagtgaattcccatcttgaagcaaatatagtttttaataattttttgataaaa 308			
DB	113	WTAWHAAWAAATATWATTTTMTTTHMAAATAATTTTMTTMAATTTAAAAATTTTAA 172			
QY	309	taacagtcaggtattatagtcacgaagcaaacataattttattgtagcaagtttaatt 368			
DB	173	AWMMMTTAAATTTTWTATWAAAMMTWAAAWAAAWAAWATATATATTATTTAA 232			
QY	369	cagaataattccaataactgatttat 395			
DB	233	AWMAAATATWMAAAAAAMWWMMWTAT 259			
RESULT	2				
CNS004ZW/c	CNS004ZW 1101 bp DNA GSS 03-JUN-1999				
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC #				
DEFINITION	BAC11E08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL055440				
VERSION	AL055440.1 GI:4932241				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 1101)				
TITLE	Genoscope.				
JOURNAL	Direct Submission				
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)				
FEATURES	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .				
	Location/Qualifiers				
	1..1101				
	/organism="Drosophila melanogaster"				
	/db_xref="taxon:7227"				
BASE COUNT	294 a 74 c 99 g 381 t 253 others				
	/clone_lib="RPCI-98"				
ORIGIN	/clone="BAC11E08"				
	/note="end : 17"				
Query Match	14.7% ; Score 61 ; DB 13 ; Length 1101 ;				



[illegible]

```

source
1. .l101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCL-98"
/clone="BACR01A24"
/note="end : TET3"

BASE COUNT      228 a   114 c   110 g   512 t   137 others
ORIGIN

Query Match          13.8% ; Score 57.2; DB 13; Length 1101;
Best Local Similarity 39.5% ; Pred. No. 0.094;
Matches 147; Conservative 56; Mismatches 165; Indels 4; Gaps 2;

Qy 43 attacatttgaactcttacgtagagaaacaactcaccaagcatttaaccatgttcataata 102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1087 AWAAWAAAAATTAAATATTTWAAAANAAAAAAAAAATTTATAAAWATWATWTWAAATAATA-A 1029

Qy 103 aatatgtacattatocgtatatatcacagtatacaaatagtagcgagaagaaatccactgt 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1028 AAAAAATAAAAATTTAATATATAAAAAATWWTAAWATAAATATTTTATAATAAATTAANAATAW 969

Qy 163 aaagcagcagggggcaccatcggtttccaagtattatataattaataattatcgtag 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 968 ATWTAAWAAAAANAAAAAAAAATAAAAANAAAAATTAATTTTAAWWTAAANWATATAA 909

Qy 223 gatgtacatgccgatagaagaagcaatttgtagatgtaaattcccactctgaaagaaa 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 908 TTAAWATWAAAAAATWAAAAATTTTAATWTTAAATTTTAAATTTTTTTTAAWTAANAAAAA 849

Qy 283 tatagtttaaat--atttattgataaataaacagtcagggtatttatagtcaccaagcaaaa 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 848 AATWAAAAANAAAAATAATWAAAAANWAAAAWATAWAAAAATAWAAAAATWAAAAAAHTAA 789

Qy 340 acataaaatttattgatgcagtttaaatcagaaatatattccaataactgattatcatcgc 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 788 ATWAAAAATTACMAAAATTYAAATTYAAAAATAAAWAAAAATTTWAWAWAWMWATAWAAAAA 729

Qy 400 tgggtacattgcc 411
    || || || || |

Db 728 AATTAAWATHCM 717

RESULT 5
CNS00238/6
LOCUS CNS00238 1101 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SFG end of BAC
            BCN01A24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL097166
VERSION AL097166.1 GI:560877
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;
         BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
         - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
         collaboration with the European Drosophila Genome Project (EDGP) -
         http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
         d'Etude du Polymorphisme Humain) with funding provided by a MRC
         project grant. The DNA was prepared from embryos by Alain Bucheton
         and Genevieve Payan. It has been constructed in the vector
         pBelOBAC11.

FEATURES             Location/Qualifiers
     source            1..1101
```

```
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db_xref="taxon:7227"  
/clone_lib="DrosBAC"  
/clone="BACN01A24"  
/note="end : sp6"
```

BASE COUNT	308 a	152 c	162 g	386 t	93 others
ORIGIN					

Query Match 13.8%; Score 57.2; DB 13; Length 1101;  
Best Local Similarity 41.4%; Pred. NO. 0.094;  
Matches 87; Conservative 38; Mismatches 85; Indels 0; Gaps 0;

Qy	186	tcaaagtattatataaattaataataattaggtagcatggccgataagaaaa	245
	:	: : : : :   :   :   :   :   :   :   :   :   :   :   :	
Db	1038	wttwtgtatwttwwaaatttatttatwttawttaatwnatatTTTTTATATAAWAA	979

Qy	246	ggcaattgtagatgtaattcccatcttgaagaataatagttttaaatattattattgata	305
		:     :     :	
Db	978	WATTTTTTTTTTAAWTTTTTTTTTTTATATTTWTTTAAAAAAATATTATWTTTAAWTTAAWTTAAAAAAW	919

[illegible]

Oy 366 attcagaaatatttcaataactgattatat 395  
:  
Db 858

RESULT 6  
CONSISTENT

LOCUS	CNS016LI	1101 bp	DNA	GSS	26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BAC16D22 of DrosBAC library from Drosophila melanogaster / fruit				

fly), genomic survey sequence.  
 ACCESSION AL106896  
 VERSION AL106896.1 GI:5624374  
 KEYWORDS GSS.

SOURCE	ORGANISM	fruit fly.
	Plasmid <i>Drosophila melanogaster</i>	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	fruit fly.	

REFERENCE  
1 (bases 1 to 1101)  
Genoscope.

**JOURNAL**  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome Project (EDGP) <http://www.edgp.ebi.ac.uk>. This *Drosophila melanogaster* BAC library (Dros BAC) was constructed from a *Drosophila* genomic library.

d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Buche and Genevieve Payan. It has been constructed in the vector

```

FEATURES
source
Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
pseudocall.

```

```
/plasmid="pBelobAC11"  
/db_xref="taxon:7227"  
/clone_lib="DrosBAC"  
/clone="BACN16D2"
```

/note="end : T7"			
203 a	220 c	84 g	158 t
BASE COUNT			436 others
ORIGIN			

Query Match 13.3%; Score 55.4; DB 13; Length 1101;  
Best Local Similarity 30.7%; Pred. No. 0.21;

[illegible]

RESULT	7				
AQ506817/c					
LOCUS	AQ506817	537 bp	DNA	GSS	29-APR-1999
DEFINITION	RPCI-11-28J17.TV RPCI-11 Homo sapiens genomic clone RPCI-11-28J17				
	, DNA sequence.				

ACCESSION AQ506817  
VERSION AQ506817.1 GI:4711564  
KEYWORDS GSS.  
SOURCE human.

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

**AUTHORS** (Pages 1 to 337)  
Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J. C.

**TITLE**  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

JOURNAL COMMENT

Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0300

Fax: 301 838 0208  
Email: [hbestigr.org](mailto:hbestigr.org)  
Clones are derived from the human BAC library RPCI-11. For BAC

laboratory availability, please contact Pieter de Jong (pieter@jeng.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([inforesgen.com](http://inforesgen.com)). BAC end search page: <http://inforesgen.com>.

```

FEATURES
    Location/Qualifiers
    Class: BAC ends.
    seq primer: T7
    http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.

```

```
1. .537
/organism="Homo sapiens"
/db_xref="GDB:76077.2"
/db_xref="GDB:76077.2"
}
}
```

```

/clone="RPCi-11-281J17"
/clone_lib="RPCi-11"
/sex="Male"

```

```
BASE COUNT      216 a    11 c    30 q    280 t
                    "RPChII Human Male BAC Library"
note=Vector: pBACC3.6; Site_1: EcoRI; Site_2: EcoRI;
/cell_type=Lymphocytes"
```



BASE COUNT	395 a	120 c	103 g	334 t	149 others
ORIGIN	/note="end : TET3"				
Query Match	13.2%; Score 54.8; DB 13; Length 1101;				
Best Local Similarity	39.8%; Pred.No.0.27;				
Matches 146; Conservative 32; Mismatches 189; Indels 0; Gaps 0;					
Qy	28	tttgaacttcaccattacatattgaaactcttacggatgagacaactcacagcatt	87		
Db	548	TTTTWAMTTTATAMATAAAAAA			
Qy	88	aatcatgttcataataaatatgtcatattacgtatatatacacogtatcaaatagtag	147		
Db	608	TTTTTTTTTTTTTTTTTTTTTAMTTTTTWTATNTTWTAAATTTTAAATAAWTTTTAT	667		
Qy	148	cgaagaataccatgtcaaagcgagcgaggccacatggttccaagtattatataaataaa	207		
Db	668	TWANNAAWATTAAAAA			
Qy	208	ttaataattcgttgaggatgtagatgcgcagataaaggaagccaatttgttagatgtaattc	267		
Db	728	TATATMTTTTWAATAATATATAAAAANATATAMTTTTTATTATTTAAAAATATWATW	787		
Qy	268	ccactctgaaagaatacagtttaaatatttatgatataaaaatacaagtcagggtattata	327		
Db	788	ATTTATATATNWNATATWTTTWTAAATTTWATMTTAAATWTMAATTAATAAWTTTATA	847		
Qy	328	gtccaagcaaaacatacaatttatgtatgcaggtttaaatccagaaatattccaataect	387		
Db	848	ATAAAAAWAAAAAATAATAAAAAATATWTATWTWATWATWTTTWAATAATWATAAW	907		
Qy	388	gattata	394		
Db	908	AAAAAA	914		

RESULT 10  
CNS00JG1/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CNS00JG1 980 bp DNA GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR38J18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
AL076232  
AL076232.1 GI:4955810  
GSS.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 980)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 9106 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> the BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and Est libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

[illegible]

## ORIGIN

[illegible]

RESULT 14  
 AU087444/c  
 LOCUS  
 DEFINITION  
 AU087444 500 bp mRNA  
 AU087444 Sugano Malaria cDNA library  
 clone XPn4669, mRNA sequence.  
 AU087444  
 AU087444.1 GI:12389585  
 EST.  
 malaria parasite P. falciparum.  
 Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; plasmodium.  
 1 (bases 1 to 500)  
 Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.  
 FULL-malaria: a database for a full-length enriched cDNA library  
 from human malaria parasite, Plasmodium falciparum  
 Nucleic Acids Res. 29 (1), 70-71 (2001)  
 20574754  
 JOURNAL  
 MEDLINE  
 COMMENT  
 Contact: Junichi Watanabe

Email: jwatanabe@manage.ims.u-tokyo.ac.jp  
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

```

a 349
FEATURES
Location/Qualifiers
1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone_lib="Sugano Malaria CDNA library"
/dev_stage="erythrocytic stage"
203 a 36 C 24 g 227 t 10 others
BASE COUNT
ORIGIN

```

Query Match	12.9%	Score 53.4;	DB 10;	Length 500;
Best Local Similarity	50.3%;	Pred. NO. 0.57;		
Matches 155;	Conservative	0;	Mismatches 152;	Indels 1;
			Gaps	1;

Qy 89 atcatgttcataataatattgtaca-ttatacgtatatatacacgtatatacaataagtag 147

499	ATAATTTAAATATAAAATAAAATAACATATAGATATATATATATATATATATATCGT	440	
QY	148	cgagaagaatccatgtcaaacgagcagggggccaccatggttccaagtattataaattataa	207
Db	439	TTAATATACACGTATAATTTTATTACGTATATCGTCCTCAATATTTTTTTTAAATATAT	380
QY	208	ttataattatggtgaggtgtacatggccatagaagaaaagcgaattgtagatggttaattc	267
Db	379	TTAGAAAATAGCTTTTATCTCAAAAATATAAAGAGAAAATTTTATTTTATATATTATACN	320
QY	268	ccatcttgaagagaatatagtttcaaatattttatgataaaatacaacagtcaggattata	327
Db	319	AATATAITTTTATATATCTATTAATATATATATACATGGATGAAGGAATAAAAAATAA	260
QY	328	gtccaacgcaaaaacataaaattttatgtatgcaagttttaaattccagaataatttccaataact	387
Db	259	ATATACATATATATATATATATATATATATATATATATAGTAAAAAGGATTGTGCTATTATA	200
QY	388	gattatat	395
Db	199	AATATTAT	192

RESULT 15	
CNS0182A	970 bp DNA
LOCUS	Drosophila melanogaster genome survey sequence 17 end of BAC
DEFINITION	BAC37C16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL108796
VERSION	ALL08796.1 GI:5629100
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Plasmid Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
	1 (bases 1 to 970)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
	- Web : www.genoscope.cns.fr)

```

FEATURES
  source
    Location/Qualifiers
      1..970
        /organism="Drosophila melanogaster"
        /plasmid="pBelobAC11"
        /db_xref="taxon:7227"
        /clone_lib="DrosBAC"
        /clone="BACN37C16"
        /note="end..773"

```

	Query Match	12.9%	Score 53.4	DB 13	Length 970
	Best Local Similarity	42.6%	Pred. No. 0.51		
	Matches 144	Conservative 13	Mismatches 181	Indels 0	Gaps 0
OY	39 caccattacatttgaaactcttcaggatgagacaactcacaagcattaatcatgttca				98
Ddb	15 CAAATTNNATANNNNNTTAANTNNNANAAAAAATAATAATMTAAM				74
OY	99 tataaatatatgtcacattatcagtatatatacacgctatacacaatagtgcgaagaatcc				158

```
Db 75 AAAAAAAAAACCCCTAATAATAATCNACAAAAATTTAAAAATAATAATAAAAAATAAAA 134
Qy 159 atgtaagcagcagggggcaccatggtttcaagttattataataattataattatg 218
   ||:|||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 ATAWAAAAAATWTTTATWAAWAAAAATATAWATAWTAATAAAAAAANTTNCTAATTTT 194
Qy 219 gtaggatgtacatgcccgcataagaaaaggcaattttagatgttaattcccatcttgaaa 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 ATTATNATAAATAANCTATAAAAAAATAAAAAATAAAAAATAAAAAACAATATAAAAA 254
Qy 279 gaaatatagtttaaatattttatgataaaaataacacagtcagggtatttagtcacagcaaa 338
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 ATAAAAAATAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 314
Qy 339 aacataaatttattgatgcaagtttaaatccagaata 376
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 AAAATAAATAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 352
```

Search completed: February 25, 2002, 17:21:06  
Job time: 16159 sec

**THIS PAGE BLANK (USPTO)**





```
misc_feature 1..234
/Note="plant DNA"
misc_feature 235..415
/Note="T-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 415; DB 6; Length 415;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcgagtttggttcacgatttgggtttgactcttcacacattacacattgaaactct 60
DB 1 gtcgagtttggttcacgatttgggtttgactcttcacacattacacattgaaactct 60
QY 61 tacgagtgagaaactcacagcattcaatcatgttcacataataatgtacattatc 120
DB 61 TACGATGAGAACAACACTCACAGCATTAATCATGTTTCATATAATATATGTACATTATAC 120
QY 121 gtatatcacgctatacaaatagtagcgaagaatccatgtaagcagcagggggcacc 180
DB 121 GTATATATACAGTATACAAATAGTAGCGAAGAAATCCATGTAAGCAGCAGGGGGCACC 180
QY 181 atggttcaagctattataataattataattatgtaggttaggtgacatggccgataa 240
DB 181 ATGTTTCAAGTATTATATAATTATAATTATAATTATGTTAGGTAGGTATACATGGCCGATAA 240
QY 241 gaaaggaattttagatgttcaattcccatcttgcgaagaatccatgtaagcagcagggggcacc 300
DB 241 GAAAGGCAATTTAGATGTTAATTCCTTCATCTTGAAGAAATATAGTTTAATATTTAT 300
QY 301 tgataaatacaagtcaggtattatagtcgaagcaaaacataaaatttattgacgaag 360
DB 301 TGATAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAACATAAAATTTATTGATGCAAG 360
QY 361 tttaaattcagaataatttcaataactgattatcatcagctggtacattgcccgtag 415
DB 361 TTTAAATTCAGAAATATTTCAATAACTGATTATATCATCGTGGTACATTGCCGTAG 415

RESULT 2
AX127757/c
LOCUS AX127757 416 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 10 from Patent WO0131042.
ACCESSION AX127757
VERSION AX127757.1 GI:14134404
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 416)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
Location/Qualifiers
source
1..416
/organism="synthetic construct"
/db_xref="taxon:32630"
/Note="3; border flanking region of elite event MS-B2"
misc_feature 1..193
misc_feature 194..416
/Note="T-DNA"
/Note="plant DNA"
BASE COUNT 137 a 72 c 54 g 152 t 1 others
ORIGIN

Query Match
Best Local Similarity 91.8%; Score 380.8; DB 6; Length 416;
Matches 404; Conservative 1; Mismatches 3; Indels 2; Gaps 2;
```

RESULT 4  
AX063413/c  
LOCUS AX063413 7599 bp DNA PAT 24-JAN-2001  
DEFINITION Sequence 5 from Patent WO0100833.  
ACCESSION AX063413  
VERSION AX063413.1 GI:12541201  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 7599)  
AUTHORS Hoffmann,B., Moillier,P. and Pelletier,G.  
TITLE Promoter expressed specifically in the cells of plant roots,  
recombinant vectors and host cells comprising same and transgenic  
plants obtained  
JOURNAL Patent: WO 0100833-A 5 04-JAN-2001;  
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)  
FEATURES Location/Qualifiers  
source l. .7599  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="ADN-T de PGK85"  
BASE COUNT 1972 a 1938 c 1937 g 1752 t  
ORIGIN  
Query Match 46.1%; Score 191.4; DB 6; Length 7599;  
Best Local Similarity 99.5%; Pred. No. 5.6e-25;  
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 223 gatgtacatgcccgcataagaaggaattgttagatgttaattcccatcttgaagaaa 282  
Db 6771 GATGTACATGTCGATAAAGAAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAA 6712  
Qy 283 tatagtttaaatattttatgataaaatacaagtcagggtattattagtcacgaagcaaaaca 342  
Db 6711 TATAGTTTAAATATTTTATTGATAAAATPACAAAGTCAGGTATTATAGTCCAAAGCAAAAACA 6652  
Qy 343 taaatttattgatgcagatttaattcagaaataatttcataaactgatttatcagctgg 402  
Db 6651 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGG 6592  
Qy 403 tacattgcccgtag 415  
Db 6591 TACATTGCCGTAG 6579  
RESULT 5  
E31991/c  
LOCUS E31991 6539 bp DNA PAT 07-FEB-2001  
DEFINITION Mutated barnase gene and transgenic plant thereof.  
ACCESSION E31991  
VERSION E31991.1 GI:13021588  
KEYWORDS JP 2000041682-A/4.  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 6539)  
AUTHORS Kazuyuki,H.F.N.N.  
TITLE Mutated barnase gene and transgenic plant thereof  
JOURNAL Patent: JP 2000041682-A 4 15-FEB-2000;  
JAPAN TOBACCO INC  
COMMENT OS Escherichia coli LE392  
PN JP 2000041682-A/4  
PD 15-FEB-2000  
PF 04-AUG-1998 JP 1998220060  
PR KAZUYUKI HAMADA,FUMIO NAKAKIDO  
PI C12N15/09,A01H5/00,C12N5/10,C12N9/22/(C12N5/10,C12R1:91), PC  
C12N15/00,  
PC C12N5/00, (C12N5/00,C12R1:91)  
CC

FEATURES Location/Qualifiers  
source l. .6539  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 1755 a 1578 c 1519 g 1687 t  
ORIGIN  
Query Match 45.5%; Score 188.8; DB 6; Length 6539;  
Best Local Similarity 99.0%; Pred. No. 1.1e-24;  
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 224 atgtacatgcccgcataagaaggaattgttagatgttaattcccatcttgaagaaaat 283  
Db 6533 ACGTACATGTCGATAGAAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAAAT 6474  
Qy 284 atagtttaaatattttatgataaaatacaagtcagggtatttatagtcacgaagcaaaacat 343  
Db 6473 ATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACAT 6414  
Qy 344 aaatttattgatgcagtttaattcagaaataatttcaataactgatttatcagctggt 403  
Db 6413 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGGT 6354  
Qy 404 acattgcccgtag 415  
Db 6353 ACATTGCCGTAG 6342  
RESULT 6  
A60109/c  
LOCUS A60109 6548 bp DNA circular PAT 22-OCT-1999  
DEFINITION Sequence 2 from Patent WO9706267.  
ACCESSION A60109  
VERSION A60109.1 GI:3715125  
KEYWORDS  
SOURCE Plasmid PTS172.  
ORGANISM Plasmids.  
REFERENCE 1 (bases 1 to 6548)  
AUTHORS De,B.M.  
TITLE Genetic transformation using a PARP inhibitor  
JOURNAL Patent: WO 9706267-A 2 20-FEB-1997;  
PLANT GENETIC SYSTEMS NV (BE)  
FEATURES Location/Qualifiers  
source l. .6548  
/organism="Plasmid PTS172"  
/db\_xref="taxon:106340"  
BASE COUNT 1756 a 1579 c 1523 g 1690 t  
ORIGIN  
Query Match 45.5%; Score 188.8; DB 6; Length 6548;  
Best Local Similarity 99.0%; Pred. No. 1.1e-24;  
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 224 atgtacatgcccgcataagaaggaattgttagatgttaattcccatcttgaagaaaat 283  
Db 6542 ACGTACATGTCGATAGAAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAAAT 6483  
Qy 284 atagtttaaatattttatgataaaatacaagtcagggtatttatagtcacgaagcaaaacat 343  
Db 6482 ATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACAT 6423  
Qy 344 aaatttattgatgcagtttaattcagaaataatttcaataactgatttatcagctggt 403  
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGGT 6363  
Qy 404 acattgcccgtag 415  
Db 6363 ACATTGCCGTAG 6342

Db	6542	ACGTACATGGTCGATACAGAAAGCAATTTGTAGTGAATTAATTCACATCTTGAAAGAAT	6483
QY	284	atagtttaaatattattattgataaaatacaagtcaggtattattagtcgaagcaaaaaacat	343
Db	6482	ATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACAT	6423
QY	344	aaatttattgatgcgaagtttaaatcagaataatttcgaactgattatcagctggt	403
Db	6422	AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGTT	6363
QY	404	acattgccgtag 415	
Db	6362	ACATTGCCGTAG 6351	
RESULT	9		
LOCUS	E31990/c		
DEFINITION	E31990 6548 bp DNA	PAT	07-FEB-2001
ACCESSION	Mutated barnase gene and transgenic plant thereof.		
VERSION	E31990		
KEYWORDS	JP 2000041682-A/3.		
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 6548)		
AUTHORS	Kazuyuki, H.F.N.N.		
TITLE	Mutated barnase gene and transgenic plant thereof		
JOURNAL	Patent: JP 2000041682-A 3 15-FEB-2000;		
COMMENT	JAPAN TOBACCO INC		
	OS Escherichia coli LE392		
	PN JP 2000041682-A/3		
	PD 15-FEB-2000		
	PF 04-AUG-1998 JP 1998220060		
	PR		
	PI KAZUYUKI HAMADA, FUMIO NAKAKITDO		
	PC C12N15/09, A01H5/00, C12N5/10, C12N9/22/(C12N5/10, C12R1:91), PC, C12N15/00,		
	CC C12N5/00, (C12N5/00, C12R1:91)		
FEATURES			
source			
FT	Key	Location/Qualifiers	
	FT source	1..6548	
		/organism="Escherichia coli LE392".	
		1..6548	
		/organism="unidentified"	
		/db_xref="taxon:32644"	
BASE COUNT	1756 a 1579 c 1523 g 1690 t		
ORIGIN			
Query Match	45.5%;	Score 188.8;	DB 6; Length 6548;
Best Local Similarity	99.0%;	Pred. No. 1.6e-24;	
Matches 190;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
QY	224	atgtacatggcgcgaataagaagcgcaattttagatgttgaattcccatcttgaagaacat	283
Db	6542	ACGTACATGGTCGATAAGCAAGCAATTTGATAGTGTAAATTCCTCATCTTGAAAGAAT	6483
QY	284	atagtttaaatattattattgataaaatacaagtcaggtattattagtcgaagcaaaaaacat	343
Db	6482	ATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACAT	6423
QY	344	aaatttattgatgcgaagtttaaatcagaataatttcgaactgattatcagctggt	403
Db	6422	AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGTT	6363
QY	404	acattgccgtag 415	
Db	6362	ACATTGCCGTAG 6351	



FEATURES  
source  
PLANT GENETIC SYSTEMS NV (BE)  
Location/Qualifiers  
1. .4946

/organism="Transformation vector pTHW107"  
/db\_xref="taxon:126810"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaagaaagcaattttagatgttaattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGGTCGATAGAAAGCAATTGTAGATGTTAATCCCATCTTGAAAGAAATAT 113  
QY 286 agtttaaatattattgataaataacaagtcagggtattatagtcacaagcaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACAAAGTCAGGTATTATAGTCCAAAGCAAAACATAA 173  
QY 346 atttattgatgcagtttaattcaattcagaataatttcaataactgattatcagctggtac 405  
|||||  
Db 174 ATTATTGATGCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGGTAC 233  
QY 406 attgcccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

RESULT 14  
LOCUS A76915 4946 bp DNA PAT 19-OCT-1999  
DEFINITION Sequence 1 from Patent EP0757102.  
ACCESSION A76915  
VERSION A76915.1 GI:6088712  
KEYWORDS Transformation vector pTHW107.  
SOURCE Transformation vector pTHW107.  
ORGANISM Transformation vector pTHW107.  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS De,B.M.  
TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR  
JOURNAL Patent: EP 0757102-A 1 05-FEB-1997;  
PLANT GENETIC SYSTEMS NV (BE)  
FEATURES Location/Qualifiers  
source  
1. .4946  
/organism="Transformation vector pTHW107"  
/db\_xref="taxon:126810"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaagaaagcaattttagatgttaattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGGTCGATAGAAAGCAATTGTAGATGTTAATCCCATCTTGAAAGAAATAT 113  
QY 286 agtttaaatattattgataaataacaagtcagggtattatagtcacaagcaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACAAAGTCAGGTATTATAGTCCAAAGCAAAACATAA 173  
QY 346 atttattgatgcagtttaattcaattcagaataatttcaataactgattatcagctggtac 405  
|||||  
Db 174 ATTATTGATGCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGGTAC 233  
QY 406 attgcccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

RESULT 15  
LOCUS AR098307  
DEFINITION Sequence 1 from patent US 6074876.  
ACCESSION AR098307  
VERSION AR098307.1 GI:12807564  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS De Block,M.  
TITLE Genetic transformation using a PARP inhibitor  
JOURNAL Patent: US 6074876-A 1 13-JUN-2000;  
FEATURES Location/Qualifiers  
source  
1. .4946  
/organism="unknown"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaagaaagcaattttagatgttaattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGGTCGATAGAAAGCAATTGTAGATGTTAATCCCATCTTGAAAGAAATAT 113  
QY 286 agtttaaatattattgataaataacaagtcagggtattatagtcacaagcaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACAAAGTCAGGTATTATAGTCCAAAGCAAAACATAA 173  
QY 346 atttattgatgcagtttaattcaattcagaataatttcaataactgattatcagctggtac 405  
|||||  
Db 174 ATTATTGATGCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGGTAC 233  
QY 406 attgcccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

Search completed: February 25, 2002, 12:51:36  
Job time: 2894 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:03:22 ; Search time 142.95 Seconds  
(without alignments)  
2488.912 Million cell updates/sec

Title: US-09-698-903B-8

Perfect score: 415

Sequence: 1 gtcgagtttggttcattga.....cagctgtacatgcgtag 415

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N:Geneseq\_1101.\*  
1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	415	100.0	415	22	Right (5') border
2	380.8	91.8	416	22	Left (3') border f
3	191.4	46.1	5864	17	Plasmid pTCol113 T-
4	191.4	46.1	5865	22	Chimeric T-DNA of
5	191.4	46.1	7599	22	Nucleotide sequenc
6	188.8	45.5	5228	22	Plasmid pT5172gelt
7	188.8	45.5	6539	21	E. coli plasmid pt
8	188.8	45.5	6548	17	Plasmid pT5174 use
9	188.8	45.5	6548	18	Plasmid pT5172. C
10	188.8	45.5	6548	21	E. coli plasmid pt
11	188.8	45.5	7492	22	Plasmid pT5346. U

c 12	188.4	45.4	1303	17	AAT39337	Plasmid pT588 (Eco
c 13	188.4	45.4	3201	12	AAQ14529	pPS029 Bt ICP codi
c 14	188.4	45.4	4832	22	AAH25423	Nucleotide sequenc
c 15	188.4	45.4	4946	18	AAT59531	T-DNA of plasmid p
c 16	188.4	45.4	4946	22	AAH25422	Nucleotide sequenc
c 17	188.4	45.4	5349	19	AAV32339	T-DNA of pT524.
c 18	188.4	45.4	5864	17	AAT39339	Plasmid pTCol113 T-
c 19	188.4	45.4	5865	22	AAQ06990	Chimeric T-DNA of
c 20	188.4	45.4	7566	14	AAQ42160	Plasmid pPS0212 co
c 21	188.4	45.4	7639	14	AAQ42159	Plasmid pJD884 con
c 22	182.4	44.0	1037	11	AAQ04705	USP-Promoter-casse
c 23	182.4	44.0	1085	11	AAQ04703	Legumin-signalpept
c 24	182.4	44.0	1160	11	AAQ04706	USP-signalpeptide
c 25	179.4	43.2	1077	22	AAH25439	Right flanking reg
c 26	177.4	42.7	3201	12	AAO15144	pVE36 Bt ICP codin
c 27	153	36.9	1186	13	AAQ25707	Chimeric neo gene
c 28	146	35.2	3153	21	AAZ29122	Plasmid DV131 comp
c 29	146	35.2	3336	21	AAZ29121	Plasmid DV130 comp
c 30	146	35.2	3694	21	AAZ29124	Plasmid DV133 used
c 31	146	35.2	3877	21	AAZ29123	Plasmid DV132 used
c 32	146	35.2	24593	6	AAH50226	Sequence of opine
c 33	146	35.2	24596	6	AAH50182	Complete nucleotid
c 34	107.6	25.9	936	22	AAF58252	Oligonucleotide D1
c 35	107.6	25.9	936	22	AAF58254	Oligonucleotide D1
c 36	107.6	25.9	936	22	AAF58257	Oligonucleotide D1
c 37	107.6	25.9	936	22	AAF58259	Oligonucleotide D2
c 38	107.6	25.9	936	22	AAF58262	Oligonucleotide D2
c 39	107.6	25.9	938	22	AAF58255	Oligonucleotide D1
c 40	106.4	25.6	936	22	AAF58252	Oligonucleotide D1
c 41	106.4	25.6	936	22	AAF58254	Oligonucleotide D1
c 42	106.4	25.6	936	22	AAF58257	Oligonucleotide D1
c 43	106.4	25.6	936	22	AAF58259	Oligonucleotide D2
c 44	106.4	25.6	936	22	AAF58262	Oligonucleotide D2
c 45	106.4	25.6	938	22	AAF58255	Oligonucleotide D1

#### ALIGNMENTS

RESULT 1  
AAD06997  
ID AAD06997 standard; DNA; 415 BP.  
AC AAD06997;  
XX  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE Right (5') border flanking region of elite event MS-B2.  
KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; ds.  
XX  
OS Chimeric - Agrobacterium sp.  
OS Chimeric - Brassica sp.  
FH  
FH Key Location/Qualifiers  
FT misc\_feature 1..234  
FT /tag= a  
FT /note= "Corresponds to plant DNA"  
FT misc\_feature 235..415  
FT /tag= b  
FT /note= "Corresponds to T-DNA"  
XX  
XX WO200131042-A2.  
XX  
XX  
PD 03-MAY-2001.  
XX  
XX 26-OCT-2000; 2000WO-EP10680.  
XX  
XX 29-OCT-1999; 99US-0430497.  
XX  
XX (AVET ) AVENTIS CROPS SCIENCE NV.



```

RESULT 3
AAAT39339/c
ID AAAT39339 standard; DNA; 5864 BP.
XX
AC AAAT39339;
XX
DT 22-JAN-1997 (first entry)
XX
DE Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.
XX
DE Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;
KW transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature complement (1..25)
FT /tag= a
FT /label= RB
FT /note= "right border of Agrobacterium T-DNA"
FT polyA_signal complement (98..330)
FT /tag= b
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7"
FT CDS complement (331..882)
FT /tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin
FT acetyltransferase"
FT complement (883..2608)
FT /tag= d
FT /label= Pssu
FT /note= "promoter of Arabidopsis Rubisco small
FT subunit gene"
FT polyA_signal complement (2659..3031)
FT /tag= e
FT /label= 3'nos
FT /note= "region containing polyA signal of nopaline
FT synthase gene of Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /tag= f
FT /label= Barnase
FT /note= "Bacillus amyloliquefaciens barnase coding
FT region"
FT promoter complement (3368..4877)
FT /tag= g
FT /label= PTA29
FT /note= "promoter of stamen-specific PTA29 gene of
FT Nicotiana tabacum"
FT promoter 4924..5216
FT /tag= h
FT /label= Pnos
FT /note= "promoter of nopaline synthase gene of
FT Agrobacterium T-DNA"
FT CDS 5217..5489
FT /tag= i
FT /label= Barstar
FT /note= "region coding for barstar of Bacillus
FT amyloliquefaciens"
FT polyA_signal 5490..5765
FT /tag= j
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT misc_feature complement (5840..5864)
FT /tag= k
FT /label= LB
FT /note= "left border of Agrobacterium T-DNA"
XX
PN W09626283-A1.
XX
PD 29-AUG-1996.

```

```

XX 21-FEB-1996; 9GWO-EP00722.
PF
XX 21-FEB-1995; 95EP-0400364.
PR
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
PA
XX
PI Botterman J, Cornelissen M, Michiels F;
XX
DR WPI; 1996-402373/40.
XX
XX Prodn. of male sterile plants by transforming with a chimaeric
PT construct - comprising a male sterility DNA e.g. barnase and a
PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
PT for generating hybrid cultivars
XX
PS Example 3; Page 33-3743-47; 56pp; English.
XX
CC Plasmid pTCO113 (AY39339) is a T-DNA vector containing a bar gene
CC under control of the PSSU promoter, a barnase gene under control
CC of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
CC gene under control of the Pnos promoter. 87% of oilseed rape
CC plants regenerated after Agrobacterium-mediated transformation
CC using pTCO113 were male sterile. Barnase expression disturbed the
CC function of stamen cells leading to male sterility. Constitutive
CC expression of barstar counteracted any low level expression of
CC barnase in non-stamen tissue.
XX
SQ Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;

Query Match 46.1%; Score 191.4; DB 17; Length 5864;
Best Local Similarity 99.5%; Pred. NO. 1.4e-30;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 223 gatgtacatggccgataaagaaagcaattttagatgttaattcccatcttgaaagaaa 282
|||||
Db 5812 GATGTACATGGTCGATAGAAAAGGCAATTTGTAGATGTTAATTCCTTCCTTGAAGAAA 5753
|||||

QY 283 tatagttaaattattattgataaaataacaagtcaggtagtattatagtcacagcaaaaaa 342
|||||
Db 5752 TATAGTTTAAATATTTATTGATAAATAACAAAGTCAGGTATTATAGTCCACGCAAAAACA 5693
|||||

QY 343 taaattattgatgcaagtttaaattcagaataattcgaactgattatcatcagctgg 402
|||||
Db 5692 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATACGCTGG 5633
|||||

QY 403 tacattgcccgtag 415
|||||
Db 5632 TACATTGCCGTAG 5620

RESULT 4
AAD06990/c
ID AAD06990 standard; DNA; 5865 BP.
XX
AC AAD06990;
XX
DT 06-AUG-2001 (first entry)
XX
DE Chimeric T-DNA of plasmid pTCO113.
XX
KW T-DNA; plasmid pTCO113; transgenic Brassica plant; transformation event;
KW male-sterility gene; chimeric; tobacco; ds.
XX
OS Chimeric - Streptomyces hygroscopicus.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Bacillus amyloliquefaciens.
OS Chimeric - Nicotiana tabacum.
OS Chimeric - Agrobacterium tumefaciens.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers

```



```
QY 403 tacattgccgtag 415
Db 5633 TACATTGCCGTAG 5621

RESULT 5
AAF25320/c
ID AAF25320 standard; DNA; 7599 BP.
XX
AC AAF25320;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plasmid PGKB5.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Synthetic.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR01768.
XX
PR 25-JUN-1999; 99FR-0008185.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Hoffmann B, Mollier P, Pelletier G;
XX
DR WPI; 2001-102893/11.
XX
CC New constitutive plant promoter active specifically in roots, useful
CC for controlling expression of pest or drought resistance genes, and
CC related transgenic plants -
XX
PS Disclosure; Fig 9; 92pp; French.
XX
CC The present sequence represents a plasmid pCKB5. The plasmid contains
CC a plant promoter that directs expression of a selected sequence in
CC root calls at all stages of development of a plant. The plant promoter
CC is used to control expression of genes in a root-specific manner,
CC especially genes that provide resistance to parasites, pests (nematodes
CC or fungi), water and salt stress, or alter sugar content or nitrogen
CC transport. Fragments of the promoter are useful as probes or primers
CC to detect or amplify at least part of the promoter.
XX
SQ Sequence 7599 BP; 1972 A; 1938 C; 1937 G; 1752 T; 0 other;

Query Match 46.1%; Score 191.4; DB 22; Length 7599;
Best Local Similarity 99.5%; Pred. No. 1.4e-30;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 223 gatgtacattgccgataagaaagcaatttttagatgttaattcccatcttgaagaaa 282
Db 6771 GATGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAA 6712
QY 283 tatagtttaaatattattgataaaataacaagtcagggtattattagtcacagcaaaaaa 342
Db 6711 TATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCAAACAAAACA 6652
QY 343 taaattattgatgcaagtttaaatcagaaatatttcaataactgattatcagctgg 402
Db 6651 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATCAGCTGG 6592
QY 403 tacattgccgtag 415
Db 6591 TACATTGCCGTAG 6579
```

```
RESULT 6
AAF86439/c
ID AAF86439 standard; DNA; 5228 BP.
XX
AC AAF86439;
XX
DT 25-JUN-2001 (first entry)
XX
DE Plasmid pTS172delta.
XX
KW Male sterile plant; RNAase inhibitor; plasmid pTS172delta; ds.
XX
OS Unidentified.
XX
PN WO200124616-A1.
XX
PD 12-APR-2001.
XX
PF 12-SEP-2000; 2000WO-JP06222.
XX
PR 30-SEP-1999; 99JP-0279307.
XX
PA (NISB ) JAPAN TOBACCO INC.
XX
PI Hamada K, Nakakido F;
XX
DR WPI; 2001-266212/27.
XX
CC Method for producing male sterile rice and maize by inserting RNAse
CC gene and RNAse inhibitor genes with promoters into the plant genome -
XX
PS Disclosure; Page 14-17; 29pp; Japanese.
XX
CC The present invention relates to a method for producing male sterile
CC plants. The method comprises inserting a promoter fragment upstream of an
CC RNAse gene and a second promoter, upstream of an RNAse inhibitor protein
CC gene and inserting it into the plant genome. The method is useful for
CC producing male sterile tobacco, lettuce and rapeseed plants, but
CC preferably rice and maize. The present sequence is a vector used in
CC the method of the present invention.
XX
SQ Sequence 5228 BP; 1384 A; 1307 C; 1263 G; 1274 T; 0 other;

Query Match 45.5%; Score 188.8; DB 22; Length 5228;
Best Local Similarity 99.0%; Pred. No. 4.6e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacattgccgataagaaagcaatttttagatgttaattcccatcttgaagaaaat 283
Db 5222 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAAT 5163
QY 284 atagtttaaatattattgataaaataacaagtcagggtattattagtcacagcaaaaaa 343
Db 5162 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCAAACAAAACA 5103
QY 344 aaatttattgatgcaagtttaaatcagaaatatttcaataactgattatcagctgg 403
Db 5102 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 5043
QY 404 acattgccgtag 415
Db 5042 ACATTGCCGTAG 5031

RESULT 7
AAZ91097/c
ID AAZ91097 standard; DNA; 6539 BP.
XX
AC AAZ91097;
XX
DT 06-JUN-2000 (first entry)
```

```

XX DE E. coli plasmid pTS431 containing mutant barnase gene.
XX KW Male sterile plant; mutant barnase gene; anther-specific expression;
XX KW low fidelity PCR; primer; plant breeding; ss.
XX OS Synthetic.
XX PN WO200008176-A1.
XX PD 17-FEB-2000.
XX PF 03-AUG-1999; 99WO-JP04167.
XX PR 04-AUG-1998; 98JP-0220060.
XX PA (NISR) JAPAN TOBACCO INC.
XX PI Hamada K, Nakakido F;
XX WPI; 2000-195581/17.
XX PT Mutate barnase gene for efficient construction of plant transformants,
XX PT particularly male sterile plants free from any undesirable characters
XX PT by specifically expressing the gene alone in anther.
XX PS Example 3; Page 23-27; 30pp; Japanese.
XX CC The invention relates to the generation of male sterile plants by
XX CC the introduction of a mutant barnase gene (AA291095) for expression
XX CC specifically in the anther of a plant. This sequence represents the
XX CC E. coli/Agrobacterium shuttle vector plasmid pTS172 which contains
XX CC the mutated barnase gene (AA291095) under control of the cauliflower
XX CC mosaic virus 35S promoter. The vector also contains a region of the
XX CC Agrobacterium T-DNA gene 7. The vector is used for transmitting the
XX CC barnase gene to plants via an Agrobacterium tumefaciens host cell.
XX CC The transformed plant is used in plant breeding.
XX SQ Sequence 6539 BP; 1755 A; 1578 C; 1519 G; 1687 T; 0 other;

Query Match 45.5%; Score 188.8; DB 21; Length 6539;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 224 atgtacatgccgtagaagaagcaatttgtagatgattcccatcttgaagaagaat 283
Db 6533 ACGTACATGGTGGTATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAGAAT 6474

Qy 284 atagtttaaatatttattgataaaatacaagtcaggtattatagtcgaagcaaaacat 343
Db 6473 ATAGTTTAAATATTATTGATAAATAACAAGTCAGGTATTATAGTCCAGCAAAACAT 6414

Qy 344 aaattattgatgaagtttaattcagaataatttcaataactgatttatcagctggt 403
Db 6413 AAATTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAAGTATTATATCAGCTGGT 6354

Qy 404 acattgccgtag 415
Db 6353 ACATTGCCGTAG 6342

RESULT 8
AAT39336/c
ID AAT39336 standard; DNA; 6548 BP.
XX AC AAT39336;
XX AC AAT39336;
XX DT 22-JAN-1997 (first entry)
XX DE Plasmid pTS174 used to obtain male sterile rice.
XX KW Plasmid pTS174; male sterile; barnase; ribonuclease; transgenic plant;

```

```

KW rice; Oryza sativa; ds; cyclic.
XX Synthetic.
XX OS
XX PH Key Location/Qualifiers
XX FT 1..2003
XX FT misc_feature /*tag= a
XX FT /label= Vector
XX FT /note= "pUC19 derived vector sequences"
XX FT complement (2019..2283)
XX FT /*tag= b
XX FT /label= 3'nos
XX FT /note= "region containing polyadenylation signal
XX FT nopaline synthase gene of Agrobacterium
XX FT T-DNA"
XX FT complement (2284..2624)
XX FT /*tag= c
XX FT /label= Barnase
XX FT /product= Bacillus amyloliquefaciens barnase
XX FT complement (2625..4313)
XX FT /*tag= d
XX FT /label= PEL
XX FT /function= promoter of the stamen-specific EI gene
XX FT 4336..5710
XX FT promoter /*tag= e
XX FT /label= P35S
XX FT /function= 35S promoter of cauliflower mosaic virus
XX FT 5711..6262
XX FT /*tag= f
XX FT /label= bar
XX FT /product= phosphinothricin acetyltransferase
XX FT 6263..6456
XX FT polyA_signal /*tag= g
XX FT /label= 3'g7
XX FT /function= region containing polyadenylation signal
XX W09626283-A1.
XX 29-AUG-1996.
XX 21-FEB-1996; 96WO-EP00722.
XX 21-FEB-1995; 95EP-0400364.
XX (PLBZ) PLANT GENETIC SYSTEMS NV.
XX Botterman J, Cornelissen M, Michiels F;
XX WPI; 1996-402373/40.
XX Prodn. of male sterile plants by transforming with a chimaeric
XX construct - comprising a male sterility DNA e.g. barnase and a
XX co-regulating gene, e.g. Barstar, into the nuclear genome, useful
XX for generating hybrid cultivars
XX Example 1; Page 33-37; 56pp; English.
XX Plasmid pTS174 (AAT39336) contains Bacillus barnase DNA under control
XX of the stamen-specific PEL promoter. Embryogenic callus from rice
XX cv. Koshihikari was transformed with pTS174 alone or with pTS88
XX (see also AAT39337), a plasmid contg. Barstar DNA under control of a
XX 35S promoter. With pTS174 alone, 1 male sterile line was recovered
XX from 48 electroporation cuvettes. With both plasmids, 7 normal
XX male sterile lines were recovered from 40 cuvettes. Barnase
XX expression disturbed the function of stamen cells leading to male
XX sterility. Constitutive expression of Barstar counteracted any low
XX level expression of barnase in non-stamen tissue.
XX SQ Sequence 6548 BP; 1757 A; 1578 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 17; Length 6548;

```



```

PF 03-AUG-1999; 99WO-JP041167.
XX
PR 04-AUG-1998; 98JP-0220060.
XX
PA (NISR ) JAPAN TOBACCO INC.
PI
XX Hamada K, Nakakido F;
XX WPI; 2000-195581/17.
XX
XX Mutate barnase gene for efficient construction of plant transformants,
PT particularly male sterile plants free from any undesirable characters
PT by specifically expressing the gene alone in anther .
XX
XX Example 3; Page 19-23; 30pp; Japanese.
XX
XX The invention relates to the generation of male sterile plants by
CC the introduction of a mutant barnase gene (AAZ91095) for expression
CC specifically in the anther of a plant. This sequence represents the
CC E. coli/Agrobacterium shuttle vector plasmid pTS172 which contains
CC the synthetic barnase gene (AAZ91094) under control of the cauliflower
CC mosaic virus 35S promoter. The vector also contains a region of the
CC Agrobacterium T-DNA gene 7. The vector is used for transmitting the
CC barnase gene to plants via an Agrobacterium tumefaciens host cell.
XX The transformed plant is used in plant breeding.
XX
SQ Sequence 6548 BP; 1756 A; 1579 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 21; Length 6548;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatggcgcgataagaagaagcgaatttgcagatgtaattcccatcttgaagaagaat 283
DB 6542 ACGTACATGGTGCATAGAAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAGAAAT 6483
QY 284 atagtttaaatatttattgataaaatacaacagtcaggtattattatcagtcggt 343
DB 6482 ATAGTTTAAATATTATTGTATTAATAACAAAGTCAGGTATTATTAGTCCCAAGCAAAAACAT 6423
QY 344 aaattattgtagcaagtttaattcagaataatttcaataactgattatcagctggt 403
DB 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAAGTATTATATCAGCTGGT 6363
QY 404 acattgcccgtag 415
DB 6362 ACATTGCCGTAG 6351

RESULT 11
AAF86441/C
ID AAF86441 standard; DNA; 7492 BP.
XX
XX AAF86441;
AC
XX
XX 25-JUN-2001 (first entry)
DT
DE Plasmid pTS346.
DE
XX Male sterile plant; RNAase inhibitor; plasmid pTS346; ds.
XX
XX Unidentified.
OS
XX WO200124616-A1.
PN
XX 12-APR-2001.
PD
XX 12-SEP-2000; 2000WO-JP06222.
PF
XX 30-SEP-1999; 99JP-0279307.
PR
XX (NISR ) JAPAN TOBACCO INC.
PA

```

---

```

XX
PI Hamada K, Nakakido F;
XX
XX WPI; 2001-266212/27.
XX
XX Method for producing male sterile rice and maize by inserting RNase
PT gene and RNase inhibitor genes with promoters into the plant genome -
XX Disclosure; Page 19-23; 29pp; Japanese.
XX
XX The present invention relates to a method for producing male sterile
CC plants. The method comprises inserting a promoter fragment upstream of an
CC RNase gene and a second promoter, upstream of an RNase inhibitor protein
CC gene and inserting it into the plant genome. The method is useful for
CC producing male sterile tobacco, lettuce and rapeseed plants, but
CC preferably rice and maize. The present sequence is a vector used in
CC the method of the present invention.
XX
SQ Sequence 7492 BP; 1987 A; 1801 C; 1752 G; 1952 T; 0 other;

Query Match 45.5%; Score 188.8; DB 22; Length 7492;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatggcgcgataagaagaagcgaatttgcagatgtaattcccatcttgaagaagaat 283
DB 7486 ACGTACATGGTGCATAGAAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAGAAAT 7427
QY 284 atagtttaaatatttattgataaaatacaacagtcaggtattattatcagtcggt 343
DB 7426 ATAGTTTAAATATTATTGTATTAATAACAAAGTCAGGTATTATTAGTCCCAAGCAAAAACAT 7367
QY 344 aaattattgtagcaagtttaattcagaataatttcaataactgattatcagctggt 403
DB 7366 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAAGTATTATATCAGCTGGT 7307
QY 404 acattgcccgtag 415
DB 7306 ACATTGCCGTAG 7295

RESULT 12
AAT39337/C
ID AAT39337 standard; DNA; 1303 BP.
XX
XX AAT39337;
AC
XX
XX 22-JAN-1997 (first entry)
DT
DE Plasmid pTS88 (EcoRI-HindIII fragment).
DE
XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH misc_feature 1..35
FT /tag= a
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
FT 36..694
FT /tag= b
FT /label= p35S
FT /function= 35S promoter of cauliflower mosaic virus
FT strain CM1841
FT CDS 695..967
FT /tag= c
FT /label= barstar
FT /product= Bacillus amyloliquefaciens barstar
FT 968..1287
FT polyA_signal
FT /tag= d

```



```

FT /label= 3'g7
FT /function= region containing polyadenylation signal
FT of gene 7 og Agrobacterium T-DNA
FT misc_feature 1288..1303
FT /*tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX
XX WO9626283-A1.
XX
XX 29-AUG-1996.
XX
XX 21-FEB-1996; 96WO-EP00722.
XX
XX 21-FEB-1995; 95EP-0400364.
XX
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX Botterman J, Cornelissen M, Michiels F;
XX
XX WPI; 1996-402373/40.
XX
XX Prodn. of male sterile plants by transforming with a chimaeric
XX construct - comprising a male sterility DNA e.g. barnase and a
XX co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX for generating hybrid cultivars
XX
XX Example 1; Page 38; 56pp; English.
XX
XX The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains
XX barstar DNA under control of a 35S promoter. The plasmid was
XX used with pTS174 (see also AAT39336) contg. barnase DNA under
XX control of the stamen-specific promoter El to produce male sterile
XX rice cv. Kochihibiki transgenic plants, and with plasmid pVEL36
XX (see also AAT39338) contg. barnase DNA under control of the stamen-
XX specific PCAS5 promoter to produce male sterile maize plants.
XX Expression of barnase (a ribonuclease) in the stamen leads to male
XX sterility. Constitutive expression of barstar counteracts possible
XX low level expression of barnase DNA in non-stamen tissue.
XX
XX Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;
XX
XX
XX Query Match 45.4%; Score 188.4; DB 17; Length 1303;
XX Best Local Similarity 99.5%; Pred. No. 5e-30;
XX Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 226 gtacatgccgataagaagcaattttagatgttaattcccatctttgaaagaaatat 285
XX 1287 GTACATGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTTGAAAGAAATAT 1228
XX
XX 286 agtttaaatattattgataaaatacaagtcaggtatttagtcccaagcaaaaacataa 345
XX 1227 AGTTTAAATATTATTGATAAAATAACAGTCAGGTATTATAGTCCAGCAAAAACATAA 1168
XX
XX 346 atttattgatgcaagttttaaatcagaataatttccaataactgatttatatcagctggtag 405
XX 1167 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATATCAGCTGGTAC 1108
XX
XX 406 attgccgtag 415
XX 1107 ATTGCCGTAG 1098
XX
XX
XX RESULT 13
XX AAQ14529/c
XX ID AAQ14529 standard; DNA; 3201 BP.
XX
XX AC AAQ14529;
XX
XX 27-JAN-1992 (first entry)
XX
XX pPS029 Bt ICP coding sequence.

```

```

XX
XX Bacillus thuringiensis; insecticidal crystal protein; ICP;
XX deletion; ss.
XX
XX Synthetic.
XX
XX WO9116432-A.
XX
XX 31-OCT-1991.
XX
XX 17-APR-1991; 91WO-EP00733.
XX
XX 18-APR-1990; 90EP-0401055.
XX
XX (PLAN-) PLANT GENETIC SYST.
XX
XX Cornelissen M, Soetaert P, Stam M, Dockx J;
XX
XX WPI; 1991-339820/46.
XX
XX Modified Bacillus thuringiensis insecticidal crystal protein
XX genes - having A and T sequences changed to G and C sequences
XX encoding same amino acids, for increased expression levels
XX
XX Disclosure; Fig 6(c); 78pp; English.
XX
XX "n" in the sequence refers to not known nucleotides.
XX pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-
XX terminal modification and the internal modification of the Bt ICP
XX coding sequence.
XX See also AAQ14529, AAQ15142-44.
XX
XX Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;
XX
XX
XX Query Match 45.4%; Score 188.4; DB 12; Length 3201;
XX Best Local Similarity 99.5%; Pred. No. 5.3e-30;
XX Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 226 gtacatgccgataagaagcaattttagatgttaattcccatcttgaagaaatat 285
XX 3141 GTACATGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTTGAAAGAAATAT 3082
XX
XX 286 agtttaaatattattgataaaatacaagtcaggtatttagtcccaagcaaaaacataa 345
XX 3081 AGTTTAAATATTATTGATAAAATAACAGTCAGGTATTATAGTCCAGCAAAAACATAA 3022
XX
XX 346 atttattgatgcaagttttaaatcagaataatttccaataactgatttatatcagctggtag 405
XX 3021 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATATCAGCTGGTAC 2962
XX
XX 406 attgccgtag 415
XX 2961 ATTGCCGTAG 2952
XX
XX
XX RESULT 14
XX AAH25423
XX ID AAH25423 standard; DNA; 4832 BP.
XX
XX AC AAH25423;
XX
XX 22-AUG-2001 (first entry)
XX
XX Nucleotide sequence of plasmid pTHW118.
XX
XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
XX fertility restorer gene; barstar gene; ss.
XX
XX Synthetic.
XX
XX Streptomyces hygroscopicus.
XX
XX Arabidopsis thaliana.
XX
XX Bacillus amyloliquefaciens.

```



```
FT CDS complement (331..882)
FT /*tag= c
FT /label= Bar
FT /product= phosphinothricin acetyltransferase
FT promoter complement (883..2608)
FT /*tag= d
FT /label= PSSU
FT /note= "promoter region of Rubisco small subunit
FT gene of Arabidopsis thaliana"
FT 3'UTR complement (2658..3031)
FT /*tag= e
FT /label= 3'nos
FT /note= "3'untranslated region contg. the poly-A
FT signal of the nopaline-synthase gene of
FT Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /*tag= f
FT /label= Barnase
FT /product= barnase
FT promoter complement (3368..4876)
FT /*tag= g
FT /label= PTA29
FT /note= "promoter region of tobacco PTA29 gene"
FT misc_RNA complement (4822..4946)
FT /*tag= h
FT /label= LB
FT /note= "T-DNA left border"
XX
XX EP757102-A1.
XX
XX
XX 05-FEB-1997.
XX
XX 04-AUG-1995; 95EP-0401844.
XX
XX 04-AUG-1995; 95EP-0401844.
XX
XX (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
XX De Block M;
XX
XX WPI; 1997-111050/11.
XX
XX Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase
XX inhibitor - reduces the cultured cells response to stress and
XX reduces metabolism
XX
XX Example 3; Page 13-16; 25pp; English.
XX
XX Plasmid pTHW107 is a vector carrying T-DNA (AAF59531) comprising a
XX barnase coding sequence under control of the tobacco TA29 gene
XX stamen-specific promoter and a phosphinothricin acetyltransferase
XX coding sequence under control of an Arabidopsis Rubisco small
XX subunit gene promoter. Oilseed rape hypocotyl explants were
XX infected with Agrobacterium tumefaciens C58C1Rif carrying vector
XX pTHW107 and helper Ti plasmid pMP60. In some treatments, the
XX hypocotyls were treated with the poly-(ADP-ribose) polymerase
XX inhibitor niacinamide (250 mg/l) 4 days prior to infection.
XX Plants regenerated from niacinamide-treated transformed calli
XX had a low copy number and displayed less variation in the
XX expression profile of the transgenes.
XX
XX Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;
```

```
Query Match 45.4%; Score 188.4; DB 18; Length 4946;
Best Local Similarity 99.5%; Pred. NO. 5.5e-30;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 226 gtacatgccgataagaaggaattgtagatttaattcccatcttgaagaataat 285
Db 54 gtacatggtcdataagaaggaattgttagatttaattcccatcttgaagaataat 113
Qy 286 agtttaaatattattgataaaataacaagtcaggtattattagtcacaagcaaaacataa 345
```

```
Db 114 agtttaaatattattgataaaataacaagtcaggtattattagtcacaagcaaaacataa 173
Qy 346 atttattgatgcaagttttaaaattcagaataatttcaataactgattatcagctggtac 405
Db 174 atttattgatgcaagttttaaaattcagaataatttcaataactgattatcagctggtac 233
Qy 406 attgccgtag 415
Db 234 attgccgtag 243
```

Search completed: February 25, 2002, 12:55:46  
Job time: 3144 sec

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:03:22 ; Search time 80.64 seconds  
(without alignments)  
1165.530 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415

Sequence: 1 gtcgagtttggtgttcatag.....cagctggtacattgcgtag 415

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/1na/PTCUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/1na/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	191.4	46.1	5864	3	US-08-894-440-4
C 2	188.8	45.5	6548	3	US-08-894-440-1
C 3	188.8	45.5	6548	3	US-08-817-188-2
C 4	188.8	45.5	7811	2	US-08-549-680A-5
C 5	188.4	45.4	1303	3	US-08-894-440-2
C 6	188.4	45.4	3200	1	US-08-453-104-23
C 7	188.4	45.4	3200	2	US-08-694-824-23
C 8	188.4	45.4	4946	3	US-08-817-188-1
C 9	188.4	45.4	5560	3	US-08-817-188-5
C 10	188.4	45.4	7566	2	US-08-232-016-23
C 11	188.4	45.4	7639	2	US-08-232-016-22
C 12	186.8	45.0	5864	3	US-08-894-440-4
C 13	177	42.7	3201	1	US-08-453-104-22
C 14	177	42.7	3201	2	US-08-694-824-22
C 15	153	36.9	1186	1	US-08-064-121-2
C 16	153	36.9	1186	1	US-08-478-015-2
C 17	153	36.9	1186	3	US-08-475-975-2
C 18	153	36.9	1186	3	US-09-084-889-2
C 19	146	35.2	3153	4	US-09-080-625-3
C 20	146	35.2	3336	4	US-09-080-625-2
C 21	146	35.2	3694	4	US-09-080-625-5
C 22	146	35.2	3877	4	US-09-080-625-4
C 23	143	34.5	24595	6	5428147-1
C 24	44.8	10.8	8654	1	US-08-920-812-6
C 25	44.8	10.8	8654	1	US-08-920-827-6
C 26	44.8	10.8	8654	1	US-08-921-177-6
C 27	44.8	10.8	8654	1	US-08-362-577C-6

28	44.8	10.8	8654	2	US-08-920-828-6	Sequence 6, Appl
C 29	44.6	10.7	5526	3	US-08-751-359-21	Sequence 21, Appl
C 30	43.6	10.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 31	43	10.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 32	42.6	10.3	8920	2	US-08-446-855A-1	Sequence 1, Appl
C 33	42.6	10.3	8920	4	US-09-150-741-1	Sequence 1, Appl
C 34	42	10.1	1316	2	US-08-871-924A-1	Sequence 1, Appl
C 35	40.4	9.7	1588	3	US-09-058-489-45	Sequence 45, Appl
C 36	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appl
C 37	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appl
C 38	39.8	9.6	24595	6	5428147-1	Patent No. 5428147
C 39	39.4	9.5	80595	4	US-09-078-294-3	Sequence 3, Appl
C 40	39.2	9.4	658	4	US-08-998-416-595	Sequence 595, App
C 41	38.8	9.3	636	4	US-08-998-416-1137	Sequence 1137, App
C 42	38.8	9.3	660	1	US-07-991-867B-32	Sequence 32, Appl
C 43	38.8	9.3	660	1	US-08-107-755A-32	Sequence 32, Appl
C 44	38.8	9.3	660	2	US-08-544-332-32	Sequence 32, Appl
C 45	38.8	9.3	1511	1	US-07-991-867B-8	Sequence 8, Appl

## ALIGNMENTS

RESULT 1

US-08-894-440-4/c

; Sequence 4, Application US/08894440

; Patent No. 6025546

; GENERAL INFORMATION:

; APPLICANT: PLANT GENETIC SYSTEMS N.V.

; TITLE OF INVENTION: Method to obtain male sterile plants

; FILE REFERENCE: NMSCOR

; CURRENT APPLICATION NUMBER: US/08/894,440

; CURRENT FILING DATE: 1997-11-12

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatencIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 5864

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of

; FEATURE:

; OTHER INFORMATION: plasmid pTCO113

; NAME/KEY: misc\_feature

; LOCATION: Complement(1)..(25))

; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((98)..(330))

; OTHER INFORMATION: region containing polyadenylation signal of gene 7

; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((331)..(882))

; OTHER INFORMATION: region coding for phosphinothricin acetyl

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((883)..(2608))

; OTHER INFORMATION: promoter of small subunit gene of Rubisco of

; OTHER INFORMATION: Arabidopsis (Pssu)

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((2659)..(3031))

; OTHER INFORMATION: region containing polyadenylation signal of

; OTHER INFORMATION: napaline synthase gene of Agrobacterium T-DNA

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: Complement((3032)..(3367))

; OTHER INFORMATION: region coding for barnase of Bacillus

; OTHER INFORMATION: anlyoliquefaciens

; FEATURE:

```

, NAME/KEY: misc_feature
, LOCATION: Complement((3368)..(4877))
, OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
, OTHER INFORMATION: tabacum (PTA29)
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (4924)..(5216)
, OTHER INFORMATION: promoter of nopaline synthase gene of
, OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (5217)..(5489)
, OTHER INFORMATION: region coding for barstar of Bacillus
, OTHER INFORMATION: amyloliquefaciens
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (5490)..(5765)
, OTHER INFORMATION: region containing polyadenylation signal of gene 7
, OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: Complement((5840)..(5864))
, OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4

```

Query Match	46.1%	Score 191.4;	DB 3;	Length 5864;
Best Local Similarity	99.5%	Pred. No. 2.7e-37;		
Matches 192;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY 223	gatgtacatgcccataagaacagccaatttggtagatgttaattcccatcttgaagaaa	282		
Db	5812 GATGTACATGTCGTATAGAAGAGCAATTGTAGATGTTAATTCOCATCTTGAAGAAA	5753		
QY 283	tatagtttaeatatttatttgataaaatacaacagtcagggtattatagtcacaagcacaataca	342		
Db	5752 TATAGTTTAAATATTATTATGATAAAATAACAAGTCAGGTATTATAGTCCAACCAAAACA	5693		
QY 343	taaattttattgatcgaagtttaaatccagaataatttccaataattcatcagctgg	402		
Db	5692 TAAATTTATTGATCGAGTTTAAATTGAGAAATATTTCATAACTGATTATATCAGCTGG	5633		
QY 403	tacattgcgctag	415		
Db	5632 TACATTGCGGTAG	5620		

```

RESULT      2
US-08-894-440-1/c
; Sequence 1, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTS174
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2003)
; OTHER INFORMATION: pUC19 derived vector sequences (vector)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2283))
; OTHER INFORMATION: 3' nos: region containing polyadenylation signal

```

```

; OTHER INFORMATION: of nopaline synthase gene of Agrobacterium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2284)..(2624))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: promoter of the stamen-specific E1 gene of rice
; OTHER INFORMATION: (PE1)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5710)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus (P35S)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5711)..(6262)
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: region containing polyadenylation signal fo gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
US-08-894-440-1

```

Query Match	45.5%	Score 188.8	DB 3	Length 6548
Best Local Similarity	99.0%	Pred. No. 1.1e-36		
Matches 190	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY 224	atgtacatgcccgtatagaagaaaggcaattttagatgtttaattcccatctctgaaagaat	283		
Db 6542	ACGTACATGGTCGATAGAAAAGGCAATTTGTAGTGTAAATCCCATCTTTGAAAGAAAT	6483		
QY 284	atagtttaaatattattattgataaaaatacaacagtcagggtattatagtcacaagcaaaaaacat	343		
Db 6482	ATAGTTTAAATATTATTATGATFAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACAT	6423		
QY 344	aaattttatgcatgcaagttttaaattcagaataatttcaataactgattcatatcagctggt	403		
Db 6422	AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAAATAACTGATTATTATCAGCTGCT	6363		
QY 404	acattgcccgtag	415		
Db 6362	ACATTGCCGTAG	6351		

```

RESULT      3
US-08-817-188-2/c
; Sequence 2, Application US/08817188
; Patent No. 6074876
;
; GENERAL INFORMATION:
;
; APPLICANT: DE BLOCK, MARC
;
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
;
; FILE REFERENCE: 2121-0127P
;
; CURRENT APPLICATION NUMBER: US/08/817,188
;
; CURRENT FILING DATE: 1997-05-15
;
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
;
; EARLIER FILING DATE: 1996-07-31
;
; EARLIER APPLICATION NUMBER: EP 95401844.6
;
; EARLIER FILING DATE: 1995-08-04
;
; NUMBER OF SEQ ID NOS: 5
;
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 2
;
; LENGTH: 6548
;
; TYPE: DNA
;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pPS172
;
; FEATURE:
;
; NAME/KEY: misc_feature

```



US-08-894-440-2/c  
; Sequence 2, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894,440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1303  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI  
; OTHER INFORMATION: fragment of pTS88  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(35)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36)..(694)  
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain  
; OTHER INFORMATION: CM1841 (P35S)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (695)..(967)  
; OTHER INFORMATION: region coding for barstar of Bacillus  
; OTHER INFORMATION: amyloliquefaciens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (968)..(1287)  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1288)..(1303)  
; OTHER INFORMATION: polylinker of pGEM2  
US-08-894-440-2

Query Match 45.4%; Score 188.4; DB 3; Length 1303;  
Best Local Similarity 99.5%; Pred. No. 1.1e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgcgataagaagaaaggcaattgttagatgttaattcccatcttgaaagaataat 285  
Db 1287 GTACATGTCGATAAGAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 1228  
QY 286 agtttaaatatttattgataaaatacaagtcagtgattattagtcacaagcaaaacataa 345  
Db 1227 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACATAA 1168  
QY 346 atttattgacgaagtttaattcagaataatttcacaactgattatcagctgggtac 405  
Db 1167 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGTAC 1108  
QY 406 attgcccgtag 415  
Db 1107 ATTGCCGTAG 1098

RESULT 6  
US-08-104-23/c  
; Sequence 23, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke

APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-23

Query Match 45.4%; Score 188.4; DB 1; Length 3200;  
Best Local Similarity 99.5%; Pred. No. 1.3e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgcgataagaagaaaggcaattgttagatgttaattcccatcttgaaagaataat 285  
Db 3141 GTACATGTCGATAAGAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 3082  
QY 286 agtttaaatatttattgataaaatacaagtcaggtattattagtcacaagcaaaacataa 345  
Db 3081 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACATAA 3022  
QY 346 atttattgacgaagtttaattcagaataatttcacaactgattatcagctgggtac 405  
Db 3021 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGTAC 2962  
QY 406 attgcccgtag 415  
Db 2961 ATTGCCGTAG 2952

RESULT 7  
US-08-694-824-23/c  
; Sequence 23, Application US/08694824  
; Patent No. 5877306



GENERAL INFORMATION:  
APPLICANT: CORNELISSEN, Marc  
APPLICANT: SOETAERT, Piet  
APPLICANT: STAM, Maïke  
APPLICANT: DOCKX, Jan  
TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
TITLE OF INVENTION: IN PLANT CELLS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/694,824  
FILING DATE: 09-AUG-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,869  
FILING DATE: 16-DEC-1992  
APPLICATION NUMBER: GB 90401055.0  
FILING DATE: 18-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Rea, Teresa S  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 010830-032  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3200 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2078..2082  
OTHER INFORMATION: /note= "Nucleotides 2078-2082  
OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 45.4%; Score 188.4; DB 2; Length 3200;  
Best Local Similarity 99.5%; Pred. No. 1.3e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 226 gtacatggccgataagaagcaatttgtagatgtaattcccatcttgaaagaaatat 285  
|||||  
Db 3141 GTACATGGTCGATAGAAAAGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 3082  
Qy 286 agtttaaatattttatgataaaataacaagtcaggtattatagtcgaagcaaaaacataa 345  
|||||  
Db 3081 AGTTTAAATATTATTATGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACATAA 3022  
Qy 346 atttatgatcgaagttaaataattcagaataatttcaataactgattatatcatcgctggtag 405  
|||||  
Db 3021 ATTTATGATGCAAGTTAAATTTCAAGAAATATTTCAATAACTGATTATATCAGCTGGTAC 2962  
Qy 406 attgcccgtag 415  
|||||  
Db 2961 ATTGCCGTAG 2952

RESULT 8  
US-08-817-188-1  
Sequence 1, Application US/08817188  
Patent No. 6074876  
GENERAL INFORMATION:  
APPLICANT: DE BLOCK, MARC  
TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
FILE REFERENCE: 2121-0127P  
CURRENT APPLICATION NUMBER: US/08/817,188  
CURRENT FILING DATE: 1997-05-15  
EARLIER APPLICATION NUMBER: PCT/EP96/03366  
EARLIER FILING DATE: 1996-07-31  
EARLIER APPLICATION NUMBER: EP 95401844.6  
EARLIER FILING DATE: 1995-08-04  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 4946  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
OTHER INFORMATION: plasmid pTHW107  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((1)..(25))  
OTHER INFORMATION: T-DNA right border (RB)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((97)..(330))  
OTHER INFORMATION: 3'g7; 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
OTHER INFORMATION: T-DNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((331)..(882))  
OTHER INFORMATION: bar; region coding for phosphinotricin acetyl  
OTHER INFORMATION: transferase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((883)..(2608))  
OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((2658)..(3031))  
OTHER INFORMATION: 3' nos; 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
OTHER INFORMATION: gene of Agrobacterium T-DNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((3032)..(3367))  
OTHER INFORMATION: barnase; region coding for barnase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((3368)..(4876))  
OTHER INFORMATION: PTA29; promoter region of TA29 gene of Nicotiana  
OTHER INFORMATION: tabacum  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((4922)..(4946))  
OTHER INFORMATION: LB: T-DNA left border  
US-08-817-188-1

Query Match 45.4%; Score 188.4; DB 3; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 1.4e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 226 gtacatggccgataagaagcaatttgtagatgtaattcccatcttgaaagaaatat 285  
|||||  
Db 54 gtacatggctcgataagaagcaatttgtagatgtaattcccatcttgaaagaaatat 113

```

QY 286 agttaaatatttattgataaaatacaagtcaggtattattagtcacaaagcaaaaacataa 345
Db 114 agttaaatatttattgataaaatacaagtcaggtattattagtcacaaagcaaaaacataa 173
QY 346 attattgatgcaggttaaatcagaataattcaataaactgattatatoscgtggtac 405
Db 174 attattgatgcaggttaaatcagaataattcaataaactgattatatoscgtggtac 233
QY 406 attgcgctag 415
Db 234 attgcgctag 243

RESULT 9
US-08-817-188-5
; Sequence 5, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pRIb6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' q7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-Ls1 gene
; NAME/KEY: misc_feature

```

```

; LOCATION: (5067)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pRIb6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; US-08-817-188-5

Query Match 45.4%; Score 188.4; DB 3; Length 5560;
Best Local Similarity 99.5%; Pred. No. 1.4e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatggccgataaagaagcaatttgatagatgttaattcccatcttgaagaataat 285
Db 41 gtacatggctgataaagaagcaatttgatagatgttaattcccatcttgaagaataat 100
QY 286 agtttaaatatttattgataaaatacaagtcaggtattattagtcacaaagcaaaaacataa 345
Db 101 agtttaaatatttattgataaaatacaagtcaggtattattagtcacaaagcaaaaacataa 160
QY 346 attattgatgcaggttaaatcagaataattcaataaactgattatatoscgtggtac 405
Db 161 attattgatgcaggttaaatcagaataattcaataaactgattatatoscgtggtac 220
QY 406 attgcgctag 415
Db 221 attgcgctag 230

```

```

RESULT 10
US-08-232-016-23/c
; Sequence 23, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016

```



```

; LOCATION: 1877..2110
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2480..3005
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3006..3665
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3666..4491
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5684..6541
; OTHER INFORMATION: /note= "Sequence complementary to
; Patent No. 5952547
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7155..7639
; OTHER INFORMATION: /note= "TR1' and TR2' promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA."
;
US-08-232-016-22

```

```

Query Match 45.4%; Score 188.4; DB 2; Length 7639;
Best Local Similarity 99.5%; Pred. No. 1.5e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatgcccataagaagaagcaattttagatgttaattcccatcttgaagaataat 285
Db 2154 GTACATGTCGATAGAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 2095

Qy 286 agtttaaatattattgataaaatacaacgtagcaggtatttagtcccaagcaaaacataa 345
Db 2094 AGTTTAAATATTATTGATATAAATAACAGTCAGGTATTATAGTCCCAAGCAAAACATAA 2035

Qy 346 atttattgacgaagtttaattcagaataattttcaataactgattatcagctggtac 405
Db 2034 ATTATTGATGCAAGTTTAAATTACAGAAATATTTCATAACTGATTATATCAGCTGTAC 1975

Qy 406 attgcccgtag 415
Db 1974 ATTCCCGTAG 1965

```

```

RESULT 12
US-08-894-440-4
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NNSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of

```

```

; OTHER INFORMATION: plasmid pTCO113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
;
US-08-894-440-4

```

```

Query Match 45.0%; Score 186.8; DB 3; Length 5864;
Best Local Similarity 98.9%; Pred. No. 3.4e-36;
Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 226 gtacatgcccataagaagaagcaattttagatgttaattcccatcttgaagaataat 285
Db 54 gtacatggtcgataagaagaagcaattttagatgttaattcccatcttgaagaataat 113

Qy 286 agtttaaatattattgataaaatacaacgtagcaggtatttagtcccaagcaaaacataa 345
Db 114 agtttaaatattattgataaaatacaacgtagcaggtatttagtcccaagcaaaacataa 173

Qy 346 atttattgacgaagtttaattcagaataattttcaataactgattatcagctggtac 405
Db 174 atttattgacgaagtttaattcagaataattttcaataactgattatcagctggtac 233

```

QY 406 attgcgctag 415  
|||||  
Db 234 attgcgctag 243

RESULT 13  
US-08-453-104-22/c  
; Sequence 22, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453.104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 42.7%; Score 177; DB 1; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 239 agaaaaggaattttagatgttaattcccatcttgaagaataatagtttaaatattt 298  
|||||  
Db 3201 AGAAAAGGCAATTCTTAGATGTTAATCCCATCTTGAAGAATAATAGTTTAAATATTT 3142  
QY 299 attgataaaatacaagtcagggtattatagttccaagcaaaaacataaaatttattgatgca 358  
|||||  
Db 3141 ATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACATAAAATTTATTGATGCA 3082

QY 359 agtttaattcagaaatatttcaataactgattatattatcagctggtgtacattgcgctag 415  
|||||  
Db 3081 AGTTTAAATTCAGAAATATTTCATAAATGATTATATATCATCAGCTGGTACATTGCGGTAG 3025

RESULT 14  
US-08-694-824-22/c  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-22

Query Match 42.7%; Score 177; DB 2; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 239 agaaaaggaattttagatgttaattcccatcttgaagaataatagtttaaatattt 298  
|||||  
Db 3201 AGAAAAGGCAATTCTTAGATGTTAATCCCATCTTGAAGAATAATAGTTTAAATATTT 3142  
QY 299 attgataaaatacaagtcagggtattatagttccaagcaaaaacataaaatttattgatgca 358  
|||||  
Db 3141 ATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACATAAAATTTATTGATGCA 3082





**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:03:22 ; Search time 1397.28 Seconds  
(without alignments)  
3191.557 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagtttggttcattga.....cagctggtacatgcgtag 415

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	em_estfun:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estom:*
5:	em_estpl:*
6:	em_estba:*
7:	em_estro:*
8:	em_estov:*
9:	em_htc:*
10:	gb_est1:*
11:	gb_est2:*
12:	gb_htc:*
13:	gb_gss:*
14:	em_gss_fun:*
15:	em_gss_hum:*
16:	em_gss_inv:*
17:	em_gss_pln:*
18:	em_gss_pro:*
19:	em_gss_rod:*
20:	em_gss_vrt:*
21:	em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query %		Length	DB	ID	Description		
		Match							
1	63	15.2	734	13	CNS010MP		AL099163 Drosophil		
C 2	61	14.7	1101	13	CNS0042W		AL055440 Drosophil		
C 3	57.2	13.8	844	13	CNS03D0I		AL238491 Tetraodon		
C 4	57.2	13.8	1101	13	CNS000B8		AL063632 Drosophil		
C 5	57.2	13.8	1101	13	CNS00238		AL097166 Drosophil		
C 6	55.4	13.3	1101	13	CNS016LI		AL106896 Drosophil		
C 7	55	13.3	537	13	AQ506817		AQ506817 RPCI-11-2		
C 8	54.8	13.2	987	13	CNS014PQ		AL104456 Drosophil		
C 9	54.8	13.2	1101	13	CNS003BD		AL064091 Drosophil		
C 10	54.6	13.2	980	13	CNS00JGI		AL076232 Drosophil		
C 11	54.6	13.2	1101	13	CNS00EVL		AL069706 Drosophil		
C 12	54.4	13.1	893	13	CNS013XE		AL103436 Drosophil		

SUMMARIES

13	54	13.0	905	13	CNS00KHX	AL077798 Drosophil
C 14	53.4	12.9	500	10	AU087444	AU087444 AU087444
C 15	53.4	12.9	970	13	CNS0182A	AL108796 Drosophil
C 16	53.2	12.8	1001	13	CNS0155H	AL105023 Drosophil
C 17	53.2	12.8	1101	13	CNS0106X	AL098595 Drosophil
C 18	53	12.8	1101	13	CNS016LI	AL106896 Drosophil
C 19	52.6	12.7	734	13	CNS010MP	AL099163 Drosophil
C 20	52.6	12.7	1013	13	CNS00J7I	AL075824 Drosophil
C 21	52.6	12.7	1101	13	CNS00EPO	AL069493 Drosophil
C 22	52.4	12.6	524	13	CNS01U9O	AL167541 Tetraodon
C 23	52.4	12.6	996	13	CNS00FUH	AL071063 Drosophil
C 24	52.2	12.6	639	13	CNS038CX	AL234458 Tetraodon
C 25	51.8	12.5	928	13	CNS00DKY	AL071865 Drosophil
C 26	51.8	12.5	1101	13	CNS0039G	AL063921 Drosophil
C 27	51.8	12.5	1101	13	CNS00FMC	AL070972 Drosophil
C 28	51.6	12.4	992	13	CNS0562R	AL322812 Tetraodon
C 29	51.6	12.4	1043	13	CNS0145P	AL103735 Drosophil
C 30	51.6	12.4	1101	13	CNS00EO7	AL069440 Drosophil
C 31	51.4	12.4	876	13	CNS009G1	AL053529 Drosophil
C 32	51.2	12.3	678	13	CNS02A0C	AL187941 Tetraodon
C 33	51.2	12.3	1001	13	CNS0140O	AL103554 Drosophil
C 34	51.2	12.3	1101	13	CNS00EVL	AL069706 Drosophil
C 35	50.6	12.2	854	11	BF274512	BF274512 GA_EB002
C 36	50.6	12.2	963	10	AL566565	AL566565
C 37	50.6	12.2	1101	13	CNS00CYH	AL060100 Drosophil
C 38	50.6	12.2	1101	13	CNS00Z2U	AL097152 Drosophil
C 39	50.4	12.1	728	13	AQ272964	AQ272964 nbxb0028P
C 40	50.4	12.1	1101	13	CNS003BB	AL064089 Drosophil
C 41	50.4	12.1	1200	13	CNS016CO	AL106578 Drosophil
C 42	50.2	12.1	681	13	B60190	B60190 CIT-HSP-200
C 43	50	12.0	563	13	AQ326762	AQ326762 nbxb0038D
C 44	50	12.0	788	13	BH126604	BH126604 BARC-Satt
C 45	50	12.0	987	13	CNS014PQ	AL104456 Drosophil

ALIGNMENTS

RESULT	1					
LOCUS	CNS010MP	734 bp	DNA	GSS	26-JUL-1999	
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC					
	BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.					
ACCESSION	AL099163	1	GI:5610774			
VERSION	AL099163.1					
KEYWORDS	GSS:					
SOURCE	fruit fly.					
ORGANISM	Plasmid Drosophila melanogaster					
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;					
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;					
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	Genoscope.					
AUTHORS	Submitted					
TITLE	Direct Submission					
JOURNAL	Drosophila melanogaster genome survey sequence					
COMMENT	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)					
	Determination of this BAC-end sequence was carried out as part of a					
	collaboration with the European Drosophila Genome Project (EDGP) -					
	http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC					
	library (Dros BAC) was made by Alain Billaud at CPH (Centre					
	d'Etude du Polymorphisme Humain) with funding provided by a MRC					
	project grant. The DNA was prepared from embryos by Alain Bucheton					
	and Genevieve Payan. It has been constructed in the vector					
	pBelobAC11.					
FEATURES	Location/Qualifiers					
source	1..734					
	/organism="Drosophila melanogaster"					
	/plasmid="pBelobAC11"					
	/db_xref="taxon:7227"					
	/clone_lib="DrosBAC"					
	/clone="BACN04L20"					



[illegible]

source	1. .1101										
	/organism="Drosophila melanogaster"										
	/db_xref="taxon:7227"										
	/clone_lib="RPCI-98"										
	/clone="BACR01A24"										
	/notes"end : TER3"										
BASE COUNT	228 a	114 c	110 g	512 t	137 others						
ORIGIN											
Query Match	13.8%; Score 57.2; DB 13; Length 1101;										
Best Local Similarity	39.5%; Pred.No. 0.094;										
Matches 147; Conservative	56; Mismatches 165; Indels 4; Gaps 2;										
Qy	43	attacatatgaaactcttacggtacgagaaacactcacaagcattaatcatgttcataata 102									
Db	1087	AWAAGAAAAATTAATATTTTWWAAAAAATAATATATAAAATWATWTTWWATAATA-A 1029									
Qy	103	aataatgtacattatcacgtatatatacacgtatatacaaaatagtagcagagaatccatgt 162									
Db	1028	AAAAAATAAATTTTAATATATAAATAATWTTAAWATAAATATTTTAAATAAATTAAGAWATAW 969									
Qy	163	aaagcagcaggggcaccatggtttccaagtattataataattataattattatggtag 222									
Db	968	ATWTAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 909									
Qy	223	gatgtacatgccgatgaagaagcaattgtgtagattgtaattccccatcttgaaagaaa 282									
Db	908	TTTAAGWATWAAAAATAWAAATAATTTAAATWTTAAATTTAAATTTTAAATAAAAAA 849									
Qy	283	tatagtttaaat---atttatgtataaataacaagtcaggctattatagttccaaagcaaaa 339									
Db	848	AATWAAAAAATAAATAATWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 789									
Qy	340	acataaatattatgtcgaagtttcaattcagaataatttcaataactgattatcctgcg 399									
Db	788	ATWAAAAATTACMAAAAYWAAATTTAAATAAATAAATAAATAAATAAATAAATAAATAA 729									
Qy	400	tggtacattgcc 411									
Db	728	AATTAAATTHCM 717									
RESULT 5											
CNS00238/c	1101 bp DNA GSS 26-JUL-1999										
LOCUS	Drosophila melanogaster genome survey sequence SP6 end of BAC										
DEFINITION	BACN01A24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.										
ACCESSION	AL097166										
VERSION	AL097166.1										
KEYWORDS	GSS.										
SOURCE	fruit fly.										
ORGANISM	Plasmid Drosophila melanogaster										
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;										
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;										
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.										
REFERENCE	1 (bases 1 to 1101)										
AUTHORS	Genoscope.										
TITLE	Direct Submission										
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;										
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr										
	- Web : www.genoscope.cns.fr)										
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pSheloBAC11.										
FEATURES	Location/Qualifiers										
source	1. .1101										

source	1. .1101										
	/organism="Drosophila melanogaster"										
	/db_xref="taxon:7227"										
	/clone_lib="RPCI-98"										
	/clone="BACR01A24"										
	/notes"end : TER3"										
BASE COUNT	228 a	114 c	110 g	512 t	137 others						
ORIGIN											
Query Match	13.8%; Score 57.2; DB 13; Length 1101;										
Best Local Similarity	39.5%; Pred.No. 0.094;										
Matches 147; Conservative	56; Mismatches 165; Indels 4; Gaps 2;										
Qy	43	attacatatgaaactcttacggtacgagaaacactcacaagcattaatcatgttcataata 102									
Db	1087	AWAAGAAAAATTAATATTTTWWAAAAAATAATATATAAAATWATWTTWWATAATA-A 1029									
Qy	103	aataatgtacattatcacgtatatatacacgtatatacaaaatagtagcagagaatccatgt 162									
Db	1028	AAAAAATAAATTTTAATATATAAATAATWTTAAWATAAATATTTTAAATAAATTAAGAWATAW 969									
Qy	163	aaagcagcaggggcaccatggtttccaagtattataataattataattattatggtag 222									
Db	968	ATWTAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 909									
Qy	223	gatgtacatgccgatgaagaagcaattgtgtagattgtaattccccatcttgaagaaa 282									
Db	908	TTTAAGWATWAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 849									
Qy	283	tatagtttaaat---atttatgtataaataacaagtcaggctatattagttccaaagcaaaa 339									
Db	848	AATWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 789									
Qy	340	acataaatatttatgcacagtttaaatctcagaataatttcaataactgattatcagc 399									
Db	788	ATWAAAAATTACMAAAAYWAAATTTAAATAAATAAATAAATAAATAAATAAATAAATAA 729									
Qy	400	tggtacattgcc 411									
Db	728	AATTAWATTTCM 717									
RESULT 5											
CNS00238/c	1101 bp DNA GSS 26-JUL-1999										
LOCUS	Drosophila melanogaster genome survey sequence SP6 end of BAC										
DEFINITION	BACN01A24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.										
ACCESSION	AL097166										
VERSION	AL097166.1										
KEYWORDS	GI:5608777										
SOURCE	GSS.										
ORGANISM	fruit fly.										
	Plasmid Drosophila melanogaster										
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;										
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;										
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.										
REFERENCE	(bases 1 to 1101)										
AUTHORS	Genoscope.										
TITLE	Direct Submission										
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;										
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr										
	- Web : www.genoscope.cns.fr)										
COMMENT	Determination of this BAC-end sequence was carried out as part of a										
	collaboration with the European Drosophila Genome Project (EDGP) -										
	http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC										
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre										
	d'Etude du Polymorphisme Humain) with funding provided by a MRC										
	project grant. The DNA was prepared from embryos by Alain Bucheton										
	and Genevieve Payan. It has been constructed in the vector										
	pSheloBAC11.										
FEATURES	Location/Qualifiers										
source	1. .1101										



**ORIGIN**

Query Match 13.3%; Score 55; DB 13; Length 537;  
Best Local Similarity 51.0%; Pred. No. 0.28;  
Matches 181; Conservative 0; Mismatches 170; Indels 4; Gaps 2;

Qy	45	tacatattgaaactcttcagcatgagaacaactcacgaagcattaatcggttgcataaaa	104
Db	453	TATATATAAATATATATAAATATATACATATATAAATATATATAAATATATATAA	394
Qy	105	tatatgtacat-tatacttatataacgcgatatacaaaatagtagcgaagaatccatgta	163
Db	393	TATATACATATATATATATATAAATATATAAATATATATAAATATATACATATA	334
Qy	164	aagcagcagggggccacct--ggttccaagtattataataattataataattatggt	220
Db	333	TATCCATATATAAATATATACATATATAAATATATATAAATATATAAATATATA	274
Qy	221	aggatgtacatggcogataaagaaaggcaatttgtagtgttaattcccattcttgaaaga	280
Db	273	AATATATATACATATATCAATATATATCAATATATATAAATATATATCAATATAA	214
Qy	281	aatatagtttaaattttatgataaaaaaacacagtcagggtattatagcccaagcaaaaa	340
Db	213	TATATAAATATATATAAATATATATAAATATATAAATATATATACATATATAAATATA	154
Qy	341	cataaattattgtcgaagtttaaatcagaaatatttcaataactgattat	395
Db	153	TATAAATATATAAATATATATAAATATATAAATATATAAATATATCAATATACATATAT	99

**RESULT 8**

CNS014PQ      CNS014PQ      987 bp      DNA      GSS      26-JUL-1999  
LOCUS      Drosophila melanogaster genome survey sequence SP6 end of BAC  
DEFINITION      BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION      AL104456  
VERSION      AL104456.1 GI:5616067  
KEYWORDS      GSS.  
SOURCE      fruit fly.  
ORGANISM      Plasmid Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE      1 (bases 1 to 987)  
AUTHORS      Genoscope.  
TITLE      Direct Submission  
JOURNAL      Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT      Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

**FEATURES**      Location/Qualifiers  
source      1..987  
          /organism="Drosophila melanogaster"  
          /plasmid="pBelOBAC11"  
          /db\_xref="taxon:7227"  
          /clone\_lib="DrosBAC"  
          /clone="BACN12P22"  
          /note="end : SP6"

BASE COUNT      257 a    122 c    122 g    241 t    245 others

**ORIGIN**

Query Match 13.2%; Score 54.8; DB 13; Length 987;

**Best Local Similarity 26.8%; Pred. No. 0.27;  
Matches 103; Conservative 100; Mismatches 182; Indels 0; Gaps 0;**

Qy	4	gagtttggttcattgatttgggttttgactcttcaccattacattatgaaactcttac	63
Db	590	RBBKBNKKKKKKKKTKTTTTTTTTCAYMWCTBKCCCCCCCCCMAMMAAAWMM	649
Qy	64	ggatgagaacaactcacaagcattatcatgtgttcataataatgtacattatacgtta	123
Db	650	AWACWAAAAATTAGAAATAAAAAAWAAAAAAAATAAATATWWAINTNTWTWWW	709
Qy	124	tatatcacgtatacaaatagtagcgaagaatccatgtcaagcagcagggggccaccatg	183
Db	710	TWTTTTTTTTTAATAAAAAAWAAAAAAWTTTAAWMTTWTTTAAAGAAWTTTATAA	769
Qy	184	gttccaagtattataataattataattatggttaggtagtcacatgcccgaagaa	243
Db	770	TTWAAATTTTTTWTWTTWATTAATTAATTAATTTWATTTWAAATTAATATWTTAN	829
Qy	244	aaggcaatttgtagtgttaattcccatcttgaaagaataatagtttaaatattattga	303
Db	830	AKWTTTWTTTTTTTTTTWTATATAAATTTTWWNAATAATWGRWTTWAATAATWAA	889
Qy	304	taaaataacaagtcaggtattatagtcacaagcaaaaacataataatttattgtagc	363
Db	890	WVAAAAAAWMDRRRAAWDWVWVAWTAATAAATTTAAAAAAAATAAAAAWVWVW	949
Qy	364	aaatccagaaatatttcaataactg	388
Db	950	WWWWTTTTTTTTTTTKAWKTKNKG	974

**RESULT 9**

CNS003BD      CNS003BD      1101 bp      DNA      GSS      03-JUN-1999  
LOCUS      Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION      BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION      AL064091  
VERSION      AL064091.1 GI:4941847  
KEYWORDS      GSS.  
SOURCE      fruit fly.  
ORGANISM      Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE      1 (bases 1 to 1101)  
AUTHORS      Genoscope.  
TITLE      Direct Submission  
JOURNAL      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT      Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

**FEATURES**      Location/Qualifiers  
source      1..1101  
          /organism="Drosophila melanogaster"  
          /db\_xref="taxon:7227"  
          /clone\_lib="RPCI-98"  
          /clone="BACR08K08"

[illegible]

[illegible]







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 25, 2002, 18:00:20 ; Search time 2331.3 Seconds  
(without alignments)  
169.833 Million cell updates/sec

Title: US-09-698-903B-9  
Perfect score: 24  
Sequence: 1 tcgaagatcagcagacctccacc 24  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	24	6	AX127756
2	24	100.0	270	6	A71431
3	24	100.0	323	6	A87282
4	24	100.0	340	6	A21284
5	24	100.0	474	1	BABARSTA
6	24	100.0	4032	6	A71435
7	24	100.0	4808	6	AR007527
8	24	100.0	4808	6	AR084093
9	24	100.0	4832	6	AX172441
10	24	100.0	5865	6	AX127748
11	24	100.0	6555	6	AR007512
12	24	100.0	6555	6	AR084078
13	20.8	86.7	273	6	A71433
14	20.8	86.7	563	6	A71436
15	20.8	86.7	5349	6	A71437
16	20.8	86.7	5611	6	A71440
17	19.8	82.5	130235	8	AC008007
18	18.2	75.8	426	9	AF062270
19	18.2	75.8	74213	2	AC025320
20	18.2	75.8	124990	9	HS434P1
21	18.2	75.8	165471	2	AC015713
22	18.2	75.8	234131	2	AC093357
23	17.8	74.2	1905	1	COXCYT5YN
24	17.8	74.2	173846	2	AC087781
25	17.8	74.2	238737	2	AC084073
26	17.6	73.3	731	9	HSDC7N08
27	17.6	73.3	3534	6	AX015400
28	17.6	73.3	4153	9	HSDC7NACTN
29	17.6	73.3	12426	9	HSDC7N1A3
30	17.6	73.3	12956	1	AE006198
31	17.6	73.3	13223	1	AVU49859
32	17.6	73.3	34646	3	U00066
33	17.6	73.3	46843	8	SPBC530
34	17.6	73.3	50502	9	HS4494016
35	17.6	73.3	52358	9	HS27C10
36	17.6	73.3	69506	8	AC012680
37	17.6	73.3	95769	8	AC013430
38	17.6	73.3	129538	8	AC008854
39	17.6	73.3	139152	8	AP002525
40	17.6	73.3	141198	2	AC023898
41	17.6	73.3	147722	2	AC016783
42	17.6	73.3	157993	2	AC073354
43	17.6	73.3	167263	2	AC063966
44	17.6	73.3	167943	2	AC026322
45	17.6	73.3	170081	2	AC055714

ALIGNMENTS

RESULT 1	AX127756	24 bp	DNA	PAT	15-MAY-2001
LOCUS	Sequence 9 from Patent WO0131042.				
DEFINITION	AX127756				
ACCESSION	AX127756.1	GI:14134403			
VERSION					
KEYWORDS	synthetic construct.				
SOURCE	synthetic construct				
ORGANISM	artificial sequence.				
REFERENCE	1 (bases 1 to 24)				
AUTHORS	Weston,B. and de Beuckeleer,M.				
TITLE	Male-sterile brassica plants and methods for producing same				
JOURNAL	Patent: WO 0131042-A 9 03-MAY-2001;				
FEATURES	Aventis CropScience N.V. (BE)				
source	Location/Qualifiers				
	l..24				
	/organism="synthetic construct"				
	/db_xref="taxon:32630"				
	/note="primer MDB8"				

BASE COUNT 7 a 9 c 4 g 4 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 24;  
Best Local Similarity 100.0%; Pred. No. 0.086;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 1 TCAGAAGTATCAGCGACCTCCACC 24

## RESULT 2

A71431 A71431 270 bp DNA PAT 07-MAY-1999  
LOCUS Sequence 1 from Patent WO9810081.  
DEFINITION A71431  
ACCESSION A71431  
VERSION A71431.1 GI:4775044  
KEYWORDS

SOURCE  
ORGANISM Bacillus amyloliquefaciens.

REFERENCE  
AUTHORS Bacteria; Firmicutes; Bacillus/Clostridium group;  
TITLE Bacillus/staphylococcus group; Bacillus.

JOURNAL  
Micheils, F. and Williams, M.  
IMPROVED BARSTAR GENE  
Patent: WO 9810081-A 1 12-MAR-1998;  
MICHIELS FRANK (BE)

FEATURES  
source  
1..270  
Location/Qualifiers

/organism="Bacillus amyloliquefaciens"

/db\_xref="taxon:1390"

1..>270

/function="INHIBITOR OF BARNASE"

/codon\_start=1

/transl\_table=11

/product="BARSTAR"

/protein\_id="CAB42577.1"

/db\_xref="GI:4775045"

/translation="MKKAVINGQIRISIDLHTLKKELALPYYGENLDALWDCLTG  
WVEYPLVLEWRQEQSKLTENGAESVLQVFRKAEGCDITILS"  
BASE COUNT 80 a 58 c 74 g 58 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 270;  
Best Local Similarity 100.0%; Pred. No. 0.095;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 32 TCAGAAGTATCAGCGACCTCCACC 55

## RESULT 3

A87282 A87282 323 bp DNA PAT 22-JAN-2000  
LOCUS Sequence 7 from Patent WO9837211.  
DEFINITION A87282  
ACCESSION A87282  
VERSION A87282.1 GI:6736047  
KEYWORDS

SOURCE  
ORGANISM unidentified.

unclassified.

1 (bases 1 to 323)

Huttner, E. and Betzner, A.S.

PROTEIN COMPLEMENTATION IN TRANSGENIC PLANTS

Patent: WO 9837211-A 7 27-AUG-1998;

GENE SHEARS PTY LTD (AU); HUTTNER ERIC (AU)

FEATURES  
source  
1..323  
Location/Qualifiers

/organism="unidentified"

## CDS

/db\_xref="taxon:32644"

1..273

/codon\_start=1

/product="BARSTAR"

/protein\_id="CAB69371.1"

/db\_xref="GI:6736048"

/translation="MKKAVINGQIRISIDLHTLKKELALPYYGENLDALWDCLTG  
WVEYPLVLEWRQEQSKLTENGAESVLQVFRKAEGCDITILS"  
BASE COUNT 98 a 69 c 87 g 69 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 323;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 32 TCAGAAGTATCAGCGACCTCCACC 55

## RESULT 4

A21284 A21284 340 bp DNA PAT 31-MAY-1994  
LOCUS Artificial barstar gene.  
DEFINITION A21284  
ACCESSION A21284  
VERSION A21284.1 GI:514151  
KEYWORDS

SOURCE  
ORGANISM unidentified.

unclassified.

1 (bases 1 to 340)

Marianti, C., Leemans, J. and De Greef, W.

Plants with modified flowers

Patent: EP 0412911-A 2 13-FEB-1991;

PLANT GENETIC SYSTEMS, N.V.

FEATURES  
source  
1..340  
Location/Qualifiers

/organism="unidentified"

/db\_xref="taxon:32644"

106 a 73 c 92 g 69 t

BASE COUNT 106 a 73 c 92 g 69 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 340;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 42 TCAGAAGTATCAGCGACCTCCACC 65

## RESULT 5

BABARSTA BABARSTA 474 bp DNA BCT 23-JUN-1996  
LOCUS Bacillus amyloliquefaciens barstar gene.  
DEFINITION X15545  
ACCESSION X15545  
VERSION X15545.1 GI:1155006  
KEYWORDS barstar; ribonuclease inhibitor.  
SOURCE Bacillus amyloliquefaciens.

ORGANISM Bacillus amyloliquefaciens.

Bacteria; Firmicutes; Bacillus/Clostridium group;

Bacillus/staphylococcus group; Bacillus.

1 (bases 1 to 436)

Hartley, R.W.

Barnase and barstar. Expression of its cloned inhibitor permits

expression of a cloned ribonuclease

J. Mol. Biol. 202 (4), 913-915 (1988)

89012012

REFERENCE 2 (bases 1 to 474)

Hartley, R.

Direct Submission

JOURNAL Submitted (14-JAN-1996) R.Hartley, LCDB/NIDDK, NIH, Bethesda, 20892  
USA, email:hartley@helix.nih.gov  
REMARK Revised by author  
COMMENT On Jan 15, 1996 this sequence version replaced gi:39311.  
See also acc# x12871.  
FEATURES  
source  
1..474  
Location/Qualifiers  
/organism="Bacillus amyloliquefaciens"  
/db\_xref="taxon:1390"  
/clones="pMT311"  
94..99  
/note="10 region"  
109..119  
/note="pot. ribosome binding site"  
124..396  
/note="barstar (AA 1 - 90)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA33551.1"  
/db\_xref="GI:39312"  
/db\_xref="SWISS-PROT:P11540"  
/translation="MKKAVINGEQIRISDLHQLKKELALPEYIGENLDALMDCLTIG  
WVEYPLVLEWRQEQSKLTENGAEVLFVREAKAEGCDITILS"  
BASE COUNT 154 a 104 c 123 g 93 t  
ORIGIN

Query Match 100.0%; Score 24; DB 1; Length 474;  
Best Local Similarity 100.0%; Pred. No. 0.098;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 155 TCAGAAGTATCAGCGACCTCCACC 178

RESULT 6  
A71435  
LOCUS A71435 4032 bp DNA circular PAT 07-MAY-1999  
DEFINITION Sequence 5 from Patent WO9810081.  
ACCESSION A71435  
VERSION A71435.1 GI:4775048  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified  
REFERENCE 1 (bases 1 to 4032)  
AUTHORS Michiels,F. and Williams,M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 5 12-MAR-1998;  
MICHELIS FRANK (BE)  
FEATURES  
source  
1..4032  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 1072 a 968 c 963 g 1029 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4032;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 3435 TCAGAAGTATCAGCGACCTCCACC 3458

RESULT 7  
AR007527/c  
LOCUS AR007527 4808 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 17 from patent US 5750867.  
ACCESSION AR007527

VERSION AR007527.1 GI:3967011  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4808)  
AUTHORS Williams,M. and Leemans,J.  
TITLE Maintenance of male-sterile plants  
JOURNAL Patent: US 5750867-A 17 12-MAY-1998;  
FEATURES  
source  
1..4808  
Location/Qualifiers  
/organism="unknown"  
BASE COUNT 1370 a 1063 c 1038 g 1333 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4808;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 8  
AR084093/c  
LOCUS AR084093 4808 bp DNA PAT 01-SEP-2000  
DEFINITION Sequence 17 from patent US 5977433.  
ACCESSION AR084093  
VERSION AR084093.1 GI:10010864  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4808)  
AUTHORS Williams,M. and Leemans,J.  
TITLE Maintenance of male-sterile plants  
JOURNAL Patent: US 5977433-A 17 02-NOV-1999;  
FEATURES  
source  
1..4808  
Location/Qualifiers  
/organism="unknown"  
BASE COUNT 1370 a 1063 c 1038 g 1333 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4808;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 9  
AX172441/c  
LOCUS AX172441 4832 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 2 from Patent WO0141558.  
ACCESSION AX172441  
VERSION AX172441.1 GI:14597553  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequence.  
REFERENCE 1 (bases 1 to 4832)  
AUTHORS de Both,G. and de Beuckeleer,M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 2 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
1..4832  
Location/Qualifiers  
/organism="synthetic construct"

misc\_feature /db\_xref="taxon:32630"  
 1883. .4065 /note="T-DNA of plasmid pTHW118"  
 BASE COUNT 1528 a 883 c 932 g 1488 t 1 others  
 ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4832;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 3222 TCAGAAGTATCAGCGACCTCCACC 3199

RESULT 10  
 AX127748  
 LOCUS AX127748 5865 bp DNA PAT 15-MAY-2001  
 DEFINITION Sequence 1 from Patent WO0131042.  
 ACCESSION AX127748  
 VERSION AX127748.1 GI:14134395  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 synthetic construct.  
 synthetic construct.  
 artificial sequence.  
 REFERENCE 1 (bases 1 to 5865)  
 AUTHORS Weston,B. and de Beuckeleer,M.  
 TITLE Male-sterile brassica plants and methods for producing same  
 JOURNAL Patent: WO 0131042-A 1 03-MAY-2001;  
 Aventis CropScience N.V. (BE)  
 FEATURES  
 source  
 1. 5865  
 Location/Qualifiers  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="T-DNA of plasmid pCol13"  
 BASE COUNT 1849 a 1095 c 1149 g 1772 t  
 ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 5865;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 5249 TCAGAAGTATCAGCGACCTCCACC 5272

RESULT 11  
 AR007512/c  
 LOCUS AR007512 6555 bp DNA PAT 04-DEC-1998  
 DEFINITION Sequence 2 from patent US 5750867.  
 ACCESSION AR007512  
 VERSION AR007512.1 GI:3966996  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Unknown.  
 Unknown.  
 Unclassified.  
 REFERENCE 1 (bases 1 to 6555)  
 AUTHORS Williams,M. and Leemans,J.  
 TITLE Maintenance of male-sterile plants  
 JOURNAL Patent: US 5750867-A 2 12-MAY-1998;  
 FEATURES  
 source  
 1. 6555  
 Location/Qualifiers  
 /organism="unknown"  
 BASE COUNT 1690 a 1611 c 1584 g 1670 t  
 ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 6555;  
 Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 993 TCAGAAGTATCAGCGACCTCCACC 970

RESULT 12  
 AR084078/c  
 LOCUS AR084078 6555 bp DNA PAT 01-SEP-2000  
 DEFINITION Sequence 2 from patent US 5977433.  
 ACCESSION AR084078  
 VERSION AR084078.1 GI:10010849  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Unknown.  
 Unknown.  
 Unclassified.

REFERENCE 1 (bases 1 to 6555)  
 AUTHORS Williams,M. and Leemans,J.  
 TITLE Maintenance of male-sterile plants  
 JOURNAL Patent: US 5977433-A 2 02-NOV-1999;  
 FEATURES  
 source  
 1. 6555  
 Location/Qualifiers  
 /organism="unknown"  
 BASE COUNT 1690 a 1611 c 1584 g 1670 t  
 ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 6555;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 993 TCAGAAGTATCAGCGACCTCCACC 970

RESULT 13  
 A71433  
 LOCUS A71433 273 bp DNA PAT 07-MAY-1999  
 DEFINITION Sequence 3 from Patent WO9810081.  
 ACCESSION A71433  
 VERSION A71433.1 GI:4775046  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 unidentified.  
 unidentified.  
 Unclassified.

REFERENCE 1 (bases 1 to 273)  
 AUTHORS Michiels,F. and Williams,M.  
 TITLE IMPROVED BARSTAR GENE  
 JOURNAL Patent: WO 9810081-A 3 12-MAR-1998;  
 FEATURES  
 source  
 1. 273  
 Location/Qualifiers  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 1. .>273  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAB42578.1"  
 /db\_xref="GI:4775047"

BASE COUNT 62 a 84 c 85 g 42 t  
 ORIGIN  
 GWVEPLVLEWRQFQSKQLTENGAEVLFQFREAKAGSCDITILS"  
 /translation="MAKKAVINGEQIRISIDLHOTLKKELALPEVYGENLDALWDCLT"

Query Match 86.7%; Score 20.8; DB 6; Length 273;  
 Best Local Similarity 91.7%; Pred. No. 4.6;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||

Db 35 TCAGGAGCATCAGCGACCTCCACC 58

RESULT 14

A71436 A71436 563 bp DNA PAT 07-MAY-1999  
LOCUS Sequence 6 from Patent WO9810081.  
DEFINITION A71436  
ACCESSION A71436  
VERSION A71436.1 GI:4775049  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 563)  
AUTHORS Michiels,F. and Williams,M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 6 12-MAR-1998;  
MICHIELS FRANK (BE)  
FEATURES  
source Location/Qualifiers  
1..563  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 152 a 133 c 145 g 133 t  
ORIGIN

Query Match 86.7%; Score 20.8; DB 6; Length 563;  
Best Local Similarity 91.7%; Pred. No. 4.8;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 37 TCAGGAGCATCAGCGACCTCCACC 60

RESULT 15

A71437 A71437 5349 bp DNA PAT 07-MAY-1999  
LOCUS Sequence 7 from Patent WO9810081.  
DEFINITION A71437  
ACCESSION A71437  
VERSION A71437.1 GI:4775050  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 5349)  
AUTHORS Michiels,F. and Williams,M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 7 12-MAR-1998;  
MICHIELS FRANK (BE)  
FEATURES  
source Location/Qualifiers  
1..5349  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 1339 a 1233 c 1290 g 1487 t  
ORIGIN

Query Match 86.7%; Score 20.8; DB 6; Length 5349;  
Best Local Similarity 91.7%; Pred. No. 5.3;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
||||| ||||| ||||| ||||| |||||  
Db 4004 TCAGGAGCATCAGCGACCTCCACC 4027

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:34 ; Search time 716.55 seconds  
(without alignments)  
28,715 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24

Sequence: 1 tcagaagatcagcgacctccacc 24

Scoring table:

IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	22	PCR primer MDB251
2	24	100.0	270	19	Wild type barstar
3	24	100.0	323	19	Barstar coding seq
4	24	100.0	340	12	Clai-HindII fragme
5	24	100.0	1303	17	Plasmid pTS88 (Eco
c 6	24	100.0	2275	22	Oligonucleotide #1
c 7	24	100.0	3544	17	PTS200 contg. p35S
c 8	24	100.0	3544	20	Nucleotide sequenc
c 9	24	100.0	4032	19	Plasmid pmV1. Sy
c 10	24	100.0	4808	15	Restriction fragme
c 11	24	100.0	4832	22	Nucleotide sequenc

12	24	100.0	4896	17	AAT08976
13	24	100.0	4896	20	AA115631
14	24	100.0	5864	17	AA139339
15	24	100.0	5865	22	AA06990
c 16	24	100.0	6555	15	AAQ53874
c 17	24	100.0	7492	22	AAF86441
18	20.8	86.7	563	19	AAV23235
19	20.8	86.7	563	19	AAV23238
20	20.8	86.7	563	19	AAV23239
21	20.8	86.7	5611	19	AAV23242
c 22	17.6	73.3	245	22	AA152182
c 23	17.6	73.3	469	22	AA139127
c 24	17.6	73.3	3534	20	AAZ41289
c 25	17.6	73.3	4721	22	AA159601
c 26	17.6	73.3	4753	22	AA157815
c 27	17.2	71.7	384	22	AA120952
c 28	17.2	71.7	384	22	AA146196
c 29	17.2	71.7	384	22	AA106663
c 30	17.2	71.7	450	22	AA111739
c 31	17.2	71.7	450	22	AA133046
c 32	17.2	71.7	450	22	AA101667
c 33	17.2	71.7	543	21	AA134918
c 34	17.2	71.7	616	21	AA150006
c 35	17.2	71.7	10223	19	AAV52206
c 36	17.2	71.7	160552	22	AA082697
c 37	17	70.8	1842	22	AA083227
c 38	16.8	70.0	3831	19	AAV52424
c 39	16.6	69.2	154	21	AAA89494
c 40	16.6	69.2	609	22	AAH67560
c 41	16.6	69.2	1239	22	AAH45076
c 42	16.6	69.2	1292	22	AA080825
c 43	16.6	69.2	1371	22	AA080826
c 44	16.6	69.2	1650	22	AAF33121
c 45	16.6	69.2	1866	17	AA141492

#### ALIGNMENTS

#### RESULT 1

AD06998	ID	AD06998	standard; DNA; 24 BP.
XX	AC	AD06998;	
XX	DT	06-AUG-2001	(first entry)
XX	DE	PCR primer MDB251	to generate the flanking region of elite event MS-B2.
XX	KW	MS-B2 elite event;	transgenic Brassica plant; transformation event;
XX	KW	male-sterility gene;	PCR primer; thermal asymmetric interlaced;
XX	KW	TAIL; ss.	
XX	OS	Bacillus amyloliquefaciens.	
XX	PN	WO200131042-A2.	
XX	PD	03-MAY-2001.	
XX	XX	26-OCT-2000;	2000WO-EF10680.
XX	XX	29-OCT-1999;	99US-0430497.
XX	XX	(AVET )	AVENTIS CROPS SCIENCE NV.
XX	PI	Weston B,	De Beuckeleer M;
XX	DR	WPI;	2001-300517/31.
XX	PT	Transgenic Brassica plants,	seeds, cells or tissues, characterized by
XX	PT	harboring specific transformation events,	particularly by presence of
XX	PT	male-sterility gene,	at specific location in its genome -

PS Example 3; Page 28; 53pp; English.

CC The present invention relates to a transgenic Brassica plant or its

CC seed, cells or tissues, characterised by harbouring a specific

CC transformation event, particularly by the presence of a male-sterility

CC gene, at a specific location in the Brassica genome. Transgenic

CC Brassica plant is useful for producing a hybrid seed by crossing the

CC transgenic plant with a male-fertile Brassica plant and harvesting the

CC hybrid seed from the transgenic Brassica plant.

CC The present sequence is primary thermal interlaced (TAIL)-PCR primer

CC MDB8 used to left (3') border flanking region of elite event MS-B2.

CC This primer corresponds to position 5249-5272 of plasmid pTC0113.

XX

SQ Sequence 24 BP; 7 A; 9 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 24; DB 22; Length 24;

Best Local Similarity 100.0%; Pred. No. 0.036;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24

DB 1 tcagaagtatcagcgacctccacc 24

|||||

RESULT 2

AAV23236

ID AAV23236 standard; DNA; 270 BP.

XX

AC AAV23236;

XX

DT 17-JUL-1998 (first entry)

XX

DE Wild type barstar DNA.

XX

KW Barstar; barnase inhibitor; fertility restoration;

XX male-sterile line; ds.

XX

OS Bacillus amyloliquefaciens.

XX

PH Key Location/Qualifiers

FT CDS 1..273

FT /\*tag= a

FT /product= barstar

FT /note= "stop codon not given"

XX

PN W09810081-A2.

XX

PD 12-MAR-1998.

XX

XX 01-SEP-1997; 97WO-EP04739.

XX

PR 03-SEP-1996; 96EP-0202446.

XX

PA (PLBZ ) PLANT GENETIC SYSTEMS NV.

XX

PI Michiels F, Williams M;

XX

DR WPI; 1998-193630/17.

XX

DR P-PSDB; AAW53344.

XX

PT DNA encoding an improved barstar protein - used to restore fertility

PT In male-sterile plant lines

XX

PS Claim 8; Pages 34-35; 54pp; English.

XX

CC The present sequence was used in the preparation of an improved

CC Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which

CC can be used to restore fertility to male-sterile lines.

CC The DNA sequence encoding the improved barstar, leads to increased

CC barstar production in tapetum cells, due to improved translation,

CC and possibly protein stability.

XX

SQ Sequence 270 BP; 80 A; 58 C; 74 G; 58 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 270;

Best Local Similarity 100.0%; Pred. No. 0.048;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24

DB 32 tcagaagtatcagcgacctccacc 55

|||||

RESULT 3

AAV60977

ID AAV60977 standard; DNA; 323 BP.

XX

AC AAV60977;

XX

DT 03-DEC-1998 (first entry)

XX

DE Barstar coding sequence.

XX

KW Barnase; barstar; IPCR; inverse polymerase chain reaction; phenotype;

XX transgenic plant; hybrid seed; male sterile plant; active enzyme;

XX regulatory protein; embryoless seed; herbicide resistance; ss.

XX

OS Synthetic.

XX

PH Key Location/Qualifiers

FT CDS 1..273

FT /\*tag= a

FT /product= "barstar"

XX

PN W09837211-A1.

XX

PD 27-AUG-1998.

XX

XX 20-FEB-1998; 98WO-GB00542.

XX

PR 21-FEB-1997; 97GB-0003681.

XX

PA (GENE-) GENE SHEARS PTY LTD.

XX

PI Betzner AS, Huttner E, Paul W, Perez P;

XX

DR WPI; 1998-467572/40.

XX

DR P-PSDB; AAW71704.

XX

PT Production of transgenic plants having a desired phenotype - by

PT using a pair of parent plants which each produce a polypeptide which

PT complement each other when crossed

XX

PS Example 1; Fig 1C; 58pp; English.

XX

CC The present invention describes a pair of parent plants for producing

CC seeds comprising: (a) a first parent plant containing at least 1 gene

CC sequence encoding a polypeptide or protein A, and (b) a second parent

CC plant containing at least 1 gene sequence encoding a polypeptide or

CC protein B; where the polypeptides A and B, when expressed in separate

CC plants, do not form an active enzyme, a regulatory protein or protein

CC which affects the functionality and/or viability and/or the structural

CC integrity of a cell, but when expressed in the same plant do form an

CC active enzyme, regulatory protein, or protein which affects the

CC structural integrity of a cell. Also described is a method for producing

CC a plant having a desired phenotype by virtue of an active enzyme, a

CC regulatory protein or a protein which affects the structural integrity

CC of a cell comprising crossing a first line with a second line where the

CC first line contains one or more gene sequences encoding a polypeptide or

CC protein A but which line does not have the desired phenotype and where

CC the second line contains one or more gene sequences encoding a

CC polypeptide or protein B which is complementary to the polypeptide or

CC protein A but which line does not have the desired phenotype. The method

CC can be used for producing plants having altered phenotypes, e.g. male-

CC sterility, embryoless seeds, altered biochemical (e.g. fatty acid)  
 CC composition or herbicide resistance. The present sequence encodes  
 CC barstar which is used in an example from the present invention.  
 XX  
 SQ Sequence 323 BP; 98 A; 69 C; 87 G; 69 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 0.049;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcagaagtatcagcgcctccacc 24  
 Db 32 tcagaagtatcagcgcctccacc 55  
 |||||

RESULT 4  
 ID AAQ10460 standard; DNA; 340 BP.  
 XX  
 AC AAQ10460;  
 DT 16-APR-1991 (first entry)  
 XX  
 DE ClaI-HindII fragment contg barstar gene.  
 XX

KW Inhibitor; extracellular ribonuclease; barnase; fertility; ss.  
 OS Bacillus amyloliquefaciens.  
 XX  
 PN EP412911-A.  
 XX  
 PD 13-FEB-1991.  
 XX  
 PF 09-AUG-1990; 90EP-0402281.  
 XX  
 PR 10-AUG-1989; 89EP-0402270.  
 XX  
 PA (PLAN-) PLANT GENETIC SYST.  
 XX  
 PI Mariani C, Leemans J, De Greef W;  
 XX  
 DR WPI; 1991-046026/07.  
 XX

Cell of fertility restored plant - in which nuclear genome in  
 transformed with foreign DNA sequence neutralising activity of  
 another prod.  
 XX  
 PS Disclosure; Fig 2; 25pp; English.  
 XX

CC The barstar gene encodes and inhibitor of barnase, which degrades  
 CC RNA molecules by hybridising the bond after a guanine residue.  
 CC The gene is used, in a chimaeric sequence, to restore fertility in  
 CC plants transformed with the gene. The promoter cassette PTA29  
 CC (EP-401194) is fused in frame with the initiating ATG.  
 XX  
 SQ Sequence 340 BP; 106 A; 73 C; 92 G; 69 T; 0 other;

Query Match 100.0%; Score 24; DB 12; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 0.05;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcagaagtatcagcgcctccacc 24  
 Db 42 tcagaagtatcagcgcctccacc 65  
 |||||

RESULT 5  
 AAT39337  
 ID AAT39337 standard; DNA; 1303 BP.  
 XX  
 AC AAT39337;

XX 22-JAN-1997 (first entry)  
 DT  
 XX Plasmid pTS88 (EcoRI-HindIII fragment).  
 DE  
 XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;  
 KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.  
 XX  
 OS Synthetic.  
 XX  
 PH Key Location/Qualifiers  
 FT misc\_feature 1..35  
 FT /tag= a  
 FT /label= pGEM2  
 FT /note= "polylinker of pGEM2"  
 FT 36..694  
 FT /tag= b  
 FT /label= P35S  
 FT /function= 35S promoter of cauliflower mosaic virus  
 FT strain CM1841  
 FT CDS 695..967  
 FT /tag= c  
 FT /label= barstar  
 FT /product= Bacillus amyloliquefaciens barstar  
 FT 968..1287  
 FT /tag= d  
 FT /label= 3'g7  
 FT /function= region containing polyadenylation signal  
 FT Of gene 7 og Agrobacterium T-DNA  
 FT misc\_feature 1288..1303  
 FT /tag= e  
 FT /label= pGEM2  
 FT /note= "polylinker of pGEM2"  
 FT  
 XX WO9626283-A1.  
 XX  
 PD 29-AUG-1996.  
 XX  
 PF 21-FEB-1996; 96WO-EP00722.  
 XX  
 PR 21-FEB-1995; 95EP-0400364.  
 XX  
 PA (PLB2 ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Botterman J, Cornelissen M, Michiels F;  
 XX  
 DR WPI; 1996-402373/40.  
 XX  
 PT Prodn. of male sterile plants by transforming with a chimaeric  
 PT construct - comprising a male sterility DNA e.g. barnase and a  
 PT co-regulating gene, e.g. barstar, into the nuclear genome, useful  
 PT for generating hybrid cultivars  
 XX  
 PS Example 1; Page 38; 56pp; English.  
 XX  
 CC The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains  
 CC barstar DNA under control of a 35S promoter. The plasmid was  
 CC used with pTS174 (see also AAT39336) contg. barnase DNA under  
 CC control of the stamen-specific promoter Ei to produce male sterile  
 CC rice cv. Kochihibiki transgenic plants, and with plasmid pVEI36  
 CC (see also AAT39338) contg. barnase DNA under control of the stamen-  
 CC specific PCA55 promoter to produce male sterile maize plants.  
 CC Expression of barnase (a ribonuclease) in the stamen leads to male  
 CC sterility. Constitutive expression of barstar counteracts possible  
 CC low level expression of barnase DNA in non-stamen tissue.  
 XX  
 SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 100.0%; Score 24; DB 17; Length 1303;  
 Best Local Similarity 100.0%; Pred. No. 0.059;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 Db 726 tcagaagtatcagcgacctccacc 749

## RESULT 6

AAF86440/c  
 ID AAF86440 standard; DNA; 2275 BP.

XX AAF86440;  
 XX

DT 25-JUN-2001 (first entry)

XX Oligonucleotide #1: SEQ ID 4.

DE Male sterile plant; RNAase inhibitor; ds.

XX Unidentified.

XX WO200124616-A1.

XX 12-APR-2001.

XX 12-SEP-2000; 2000WO-JP06222.

XX 30-SEP-1999; 99JP-0279307.

XX (NTSB ) JAPAN TOBACCO INC.

XX Hamada K, Nakakido F;

XX WPI; 2001-266212/27.

XX Method for producing male sterile rice and maize by inserting RNAse  
 gene and RNAse inhibitor genes with promoters into the plant genome -

XX Disclosure; Page 17-19; 29pp; Japanese.

XX The present invention relates to a method for producing male sterile  
 plants. The method comprises inserting a promoter fragment upstream of an  
 RNAse gene and a second promoter, upstream of an RNAse inhibitor protein  
 gene and inserting it into the plant genome. The method is useful for  
 producing male sterile tobacco, lettuce and rapeseed plants, but  
 preferably rice and maize. The present sequence is an oligonucleotide  
 used in the method of the present invention.

XX Sequence 2275 BP; 604 A; 496 C; 496 G; 679 T; 0 other;

Query Match 100.0%; Score 24; DB 22; Length 2275;  
 Best Local Similarity 100.0%; Pred. No. 0.063;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 Db 533 TCAGAAGTATCAGCGACCTCCACC 510

## RESULT 7

AAAT17246/c  
 ID AAAT17246 standard; DNA; 3544 BP.

XX AAAT17246;

DT 12-AUG-1996 (first entry)

XX P35S-bar-3'nos and PCA55-barstar-3'nos.

XX Cl; Cl-S; transgenic plant; male sterility; colour-linked restorer;  
 anthocyanin; aleurone; B-peru; bar; barstar; bar\*;

XX Synthetic.

XX

Key Location/Qualifiers  
 misc\_signal 3227..3504  
 /tag= a  
 /label= 3'nos  
 /note= "3' regulatory sequence contg. the  
 polyadenylation signal of the nopaline  
 synthase gene of Agrobacterium T-DNA"  
 CDS 2675..3226  
 /tag= b  
 /label= bar  
 /note= "coding region of bar gene of  
 Streptomyces hygroscopicus"  
 promoter 1841..2674  
 /tag= c  
 /label= P35S  
 /note= "35S promoter of Cauliflower Mosaic Virus"  
 promoter complement (626..1803)  
 /tag= d  
 /label= PCA55  
 /note= "promoter of CA55 gene of Zea mays"  
 complement (353..625)  
 /tag= e  
 /label= barstar  
 /note= "coding region of barstar gene of  
 Bacillus amyloliquefaciens"  
 complement (30..352)  
 /tag= f  
 /label= 3'nos  
 /note= "3' regulatory sequence contg. the  
 polyadenylation signal of the nopaline  
 synthase gene of Agrobacterium T-DNA"  
 W059334634-A2.  
 21-DEC-1995.  
 06-JUN-1995; 95WO-EP02157.  
 06-JUN-1994; 94US-0254776.  
 (PLBZ ) PLANT GENETIC SYSTEMS NV.  
 Krebbers E, Leemans J, Williams M;  
 WPI; 1996-049664/05.  
 Transgenic plants contg. male sterility and colour-linked restorer  
 genotypes - used for prodn. of male sterile seeds identifiable from  
 their colour, also new truncated anthocyanin regulatory genes and  
 aleurone specific promoters  
 Disclosure; Page 72-74; 104pp; English.  
 Shortened forms of anthocyanin (An) regulatory genes (ARG) are easier  
 to manipulate than the complete gene and still provide An prodn.  
 The Cl gene (and the Cl-S gene) can be considerably shortened  
 while still retaining, under appropriate conditions, its  
 capability of conditioning anthocyanin prodn. in the aleurone  
 of seeds of cereal plants such as corn. A pref. shortened Cl gene  
 is comprised in pCOL9 (see AAT08975). The full Cl gene sequence is  
 given in AAT08973.  
 A truncated B-peru gene (presumed sequence = AAT08674; actual  
 sequence = AAT08977) may also be used.  
 Plasmids pTS256 (AAT08976) and pTS200 (AAT17246) were used in the  
 construction of vectors comprising the Cl and B-peru genes  
 as well as male-sterility gene and a selectable marker gene.  
 Sequence 3544 BP; 1011 A; 847 C; 767 G; 919 T; 0 other;

Query Match 100.0%; Score 24; DB 17; Length 3544;  
 Best Local Similarity 100.0%; Pred. No. 0.067;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 ||||||||||||||||||  
 Db 594 TCAGAAGTATCAGCGACCTCCACC 571

RESULT 8  
 AAX15632/c  
 ID AAX15632 standard; DNA; 3544 BP.  
 XX AC AAX15632;  
 XX DT 07-MAY-1999 (first entry)  
 XX DE Nucleotide sequence of an EcoRI-HindIII fragment of plasmid pTS256.  
 XX KW C1 gene; maize; male-sterile corn line; anthocyanin production;  
 XX KW pTS256; ds.  
 XX OS Synthetic.  
 XX PN US5880331-A.  
 XX PD 09-MAR-1999.  
 XX PF 07-JUN-1995; 95US-0485139.  
 XX PR 07-JUN-1995; 95US-0485139.  
 XX PR 06-JUN-1994; 94US-0254776.  
 XX PA (PUBZ ) PLANT GENETIC SYSTEMS NV.  
 XX PI Krebbers E, Leemans J, Williams M;  
 XX WPI: 1999-204053/17.  
 XX PT Process for maintaining a male-sterile corn line - using male  
 PT sterile plants lacking functional regulatory gene for anthocyanin  
 PT production  
 XX PS Example 2; Columns 37-40; 35pp; English.  
 XX CC The present sequence represents the nucleotide sequence of an  
 CC EcoRI-hindIII fragment of plasmid pTS256, comprising the chimeric  
 CC gene PCA55-barstar-3'nos. It is used in the course of the invention. The  
 CC specification describes a process for maintaining a male-sterile corn  
 CC line, using male sterile parent plants lacking a functional gene for  
 CC anthocyanin production, and a maintainer corn line comprising male  
 CC fertile parent plants containing foreign DNA comprising a restorer gene  
 CC and an active regulatory protein gene. By using the anthocyanin gene,  
 CC the colour of the male-sterile plants will differ from that of the  
 CC male fertile plants. This will enable the seeds harvested from the  
 CC plants to be easily separated into those that will grow into  
 CC male-fertile plants and those that will grow into male-sterile plants.  
 XX SQ Sequence 3544 BP; 1011 A; 847 C; 767 G; 919 T; 0 other;

Query Match 100.0%; Score 24; DB 20; Length 3544;  
 Best Local Similarity 100.0%; Pred. No. 0.067;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 ||||||||||||||||||  
 Db 594 TCAGAAGTATCAGCGACCTCCACC 571

RESULT 9  
 AAV23237  
 ID AAV23237 standard; DNA; 4032 BP.  
 XX AC AAV23237;  
 XX KW

DT 17-JUL-1998 (first entry)  
 XX Plasmid pMW71.  
 XX KW Barstar; barnase inhibitor; fertility restoration;  
 KW male-sterile line; plasmid pMW71; circular; ds.  
 XX OS Synthetic.  
 XX FH Key Location/Qualifiers  
 XX FT promoter /\*tag= a  
 FT /note= "promoter region of rice actin gene -  
 FT contains an intron in the leader"  
 FT CDS 3401..3676  
 FT /\*tag= b  
 FT /product= barstar  
 FT 3'UTR 3677..4003  
 FT /\*tag= c  
 FT /note= "region containing 3' untranslated end of  
 FT the nopaline synthase gene of Agrobacterium  
 FT T-DNA"  
 XX PN WO9810081-A2.  
 XX PD 12-MAR-1998.  
 XX PF 01-SEP-1997; 97WO-EP04739.  
 XX PR 03-SEP-1996; 96EP-0202446.  
 XX PA (PLEBZ ) PLANT GENETIC SYSTEMS NV.  
 XX PI Michiels F, Williams M;  
 XX WPI: 1998-193630/17.  
 XX PT DNA encoding an improved barstar protein - used to restore fertility  
 PT in male-sterile plant lines  
 XX PS Example 3; Pages 37-39; 54pp; English.  
 XX CC The present sequence was used in the preparation of an improved  
 CC Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which  
 CC can be used to restore fertility to male-sterile lines.  
 CC The DNA sequence encoding the improved barstar, leads to increased  
 CC barstar production in tapetum cells, due to improved translation,  
 CC and possibly protein stability.  
 XX SQ Sequence 4032 BP; 1072 A; 968 C; 963 G; 1029 T; 0 other;  
 Query Match 100.0%; Score 24; DB 19; Length 4032;  
 Best Local Similarity 100.0%; Pred. No. 0.068;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 tcagaagtatcagcgacctccacc 24  
 ||||||||||||||||||  
 Db 3435 tcagaagtatcagcgacctccacc 3458  
 RESULT 10  
 AAQ53889/c  
 ID AAQ53889 standard; DNA; 4808 BP.  
 XX AC AAQ53889;  
 XX DT 27-JUN-1994 (first entry)  
 XX DE Restriction fragment of construct carrying plant maintainer gene.  
 XX KW Maintainer gene; sterile; sterility; homogenous population; hybrid;  
 KW seed; fertility restorer gene; pollen lethality gene; ss.

```

XX OS Synthetic.
XX FH Key Location/Qualifiers
FT FT misc_signal complement (18..401)
FT FT /*tag= a
FT FT /label= 3' nos.
FT FT /note= "3' regulatory sequence containing the
FT FT polyadenylation site derived from
FT FT Agrobacterium T-DNA nopaline synthase gene."
FT FT CDS complement (402..737)
FT FT /*tag= b
FT FT /label= Barnase.
FT FT /note= "Coding region of the barnase gene of
FT FT Bacillus amyloliquefaciens."
FT FT promoter complement (738..1944)
FT FT /*tag= c
FT FT /label= P2M13.
FT FT /note= "Promoter region of the Zm 13 gene of Zea
FT FT maysa."
FT FT misc_signal complement (1945..2281)
FT FT /*tag= d
FT FT /label= 3' nos.
FT FT complement (2282..2554)
FT FT /*tag= e
FT FT /label= Barstar.
FT FT /note= "Coding region of the barstar gene of
FT FT Bacillus amyloliquefaciens."
FT FT promoter complement (2555..3099)
FT FT /*tag= f
FT FT /label= PTA29.
FT FT /note= "Promoter region of the TA29 gene of
FT FT Nicotiana tabacum."
FT FT promoter 3100..3932
FT FT /*tag= g
FT FT /label= 3S3.
FT FT /note= "3S3" promoter sequence derived from
FT FT cauliflower mosaic virus isolate Cabbb-JI."
FT FT CDS 3933..4484
FT FT /*tag= h
FT FT /note= "Coding region of the phosphinothricin
FT FT acetyltransferase gene."
FT FT misc_signal 4485..4763
FT FT /*tag= i
FT FT /label= 3' nos.
FT FT
XX WO325695-A.
XX
XX 23-DEC-1993.
XX
XX 11-JUN-1993; 93WO-EP01489.
XX
XX 12-JUN-1992; 92US-0899072.
XX 03-NOV-1992; 92US-0970840.
XX
XX (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
XX Leemans J, Williams M;
XX
XX WPI; 1994-007552/01.
XX
XX Maintainer gene for maintenance of male-sterile plants -
XX comprises fertility-restorer gene and pollen-lethality gene
XX
XX Example 2; Page 73-75; 87pp; English.
XX
XX A maintainer gene of plants, pref. a foreign chimeric gene,
XX comprises (a) a fertility restorer gene which comprises a fertility
XX restorer DNA and (ii) a restorer promoter capable of
XX directing the expression of the fertility restorer DNA and (b) a
XX pollen lethality gene that is selectively expressed in microspores
XX and/or pollen of the plant to prevent the production of functional
XX pollen and which comprises (i) a pollen lethality DNA and (ii) a

```

```

CC pollen specific promoter capable of directing expression of the
CC pollen lethality DNA. Plants transformed with this DNA (maintainer
CC plants) can be used to maintain a homogenous population of male
CC sterile plants for the production of hybrid seed. This sequence is
CC an EcoRI-HindIII restriction fragment of the construct designated
CC pTS218 and comprises the maintainer gene described.
XX
SQ Sequence 4808 BP; 1370 A; 1064 C; 1037 G; 1333 T; 4 other;

Query Match 100.0%; Score 24; DB 15; Length 4808;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24
|||||
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 11
AAH25423/c
ID AAH25423 standard; DNA; 4832 BP.
XX
XX AC AAH25423;
XX
XX DT 22-AUG-2001 (first entry)
XX
XX DE Nucleotide sequence of plasmid pTHW118.
XX
XX KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
XX fertility restorer gene; barstar gene; ss.
XX
XX OS Synthetic.
XX OS Streptomyces hygroscopicus.
XX OS Arabidopsis thaliana.
XX OS Bacillus amyloliquefaciens.
XX OS Nicotiana tabacum.
XX
XX FH Key Location/Qualifiers
FT FT misc_feature 1..25
FT FT /*tag= a
FT FT /note= "right border repeat from TL-DNA from pTIB6S3"
FT FT misc_feature 26..53
FT FT /*tag= b
FT FT /note= "synthetic polylinker derived sequences"
FT FT misc_feature 54..90
FT FT /*tag= c
FT FT /note= "residual sequence from TL-DNA at right
FT FT border repeat"
FT FT misc_feature 91..97
FT FT /*tag= d
FT FT /note= "synthetic polylinker derived sequences"
FT FT 3'UTR complement (98..309)
FT FT /*tag= e
FT FT /note= "3' UTR from TL-DNA gene 7 of pTIB6S3"
FT FT misc_feature 310..330
FT FT /*tag= f
FT FT /note= "synthetic polylinker derived sequences"
FT FT CDS complement (331..882)
FT FT /*tag= g
FT FT /note= "Streptomyces hygroscopicus bialaphos
FT FT resistance (bar) gene"
FT FT promoter complement (883..2608)
FT FT /*tag= h
FT FT /note= "atSIA ribulose-1,5-biphosphate carboxylase
FT FT small subunit gene from Arabidopsis thaliana"
FT FT misc_feature 2609..2658
FT FT /*tag= i
FT FT /note= "synthetic polylinker derived sequences"
FT FT 3'UTR complement (2659..2919)
FT FT /*tag= j
FT FT /note= "taql fragment from 3' UTR of nopaline
FT FT synthase gene from T-DNA of pTIF37 and

```

FT misc\_feature 2920..2940 containing plant polyadenylation signals"  
 FT /\*tag= k  
 FT /note= "synthetic polylinker derived sequences"  
 FT 2941..2980  
 FT 3'UTR /\*tag= l  
 FT /note= "downstream of Bacillus amyloliquefaciens  
 FT barstar coding region"  
 FT CDS complement (2981..3253)  
 FT /\*tag= m  
 FT /note= "Barstar gene coding region from Bacillus  
 FT amyloliquefaciens"  
 FT promoter complement (3254..4762)  
 FT /\*tag= n  
 FT /note= "anther-specific gene TA29 promoter from  
 FT Nicotiana tabacum"  
 FT misc\_feature 4763..4807  
 FT /\*tag= o  
 FT /note= "synthetic polylinker derived sequences"  
 FT misc\_feature 4808..4832  
 FT /\*tag= p  
 FT /note= "left border repeat from TL-DNA from pTib6S3"  
 FT WO200141558-A1.  
 PN  
 XX  
 XX  
 PD 14-JUN-2001.  
 XX  
 XX  
 PF 06-DEC-2000; 2000WO-EP12872.  
 XX  
 PF 08-DEC-1999; 99US-0457037.  
 XX  
 PR (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 PA De Both G, De Beuckeleer M;  
 XX  
 PI WPI; 2001-381419/40.  
 XX  
 DR Transgenic winter oilseed rape plants suited for producing hybrid seed  
 XX with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome -  
 XX  
 XX Example 1; Page 80-82; 98pp; English.  
 XX  
 CC The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene, and the other  
 CC plant has an expression cassette comprising a fertility restorer gene,  
 CC integrated into the genome. The fertility restorer gene is capable of  
 CC preventing the activity of the male-sterility gene. The plant pair is  
 CC useful for producing hybrid seed. Plants developed from the hybrid  
 CC seed have agronomic performance, genetic stability and adaptability to  
 CC different genetic backgrounds. The present sequence represents  
 CC plasmid pTHW118. This plasmid comprises the barstar gene, which acts as  
 CC a fertility restorer gene. The plasmid is used to create transgenic  
 CC plants of the invention.  
 XX  
 SQ Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;  
 XX  
 XX Query Match 100.0%; Score 24; DB 22; Length 4832;  
 XX Best Local Similarity 100.0%; Pred. No. 0.069;  
 XX Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 tcagaagtatcagcagctccacc 24  
 ||||||||||||||||||  
 Db 3222 TCAGAAGTATCAGCAGCTCCACC 3199  
 RESULT 12  
 AAT08976  
 ID AAT08976 standard; DNA; 4896 BP.  
 XX  
 AC AAT08976;

XX 12-AUG-1996 (first entry)  
 DT  
 XX  
 XX PTS256 contg. P35S-bar-3'nos and PTA29-barstar-3'nos.  
 DE  
 XX C1; C1-S; transgenic plant; male sterility; colour-linked restorer;  
 KW anthocyanin; aleurone; B-peru; bar; barstar; bar\*; ds.  
 XX  
 XX Synthetic.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH misc\_signal complement (39..317)  
 FT /\*tag= a  
 FT /label= 3'nos  
 FT /note= "3' regulatory sequence contg. the  
 FT polyadenylation signal of the nopaline  
 FT synthase gene of Agrobacterium T-DNA"  
 FT CDS complement (318..869)  
 FT /\*tag= b  
 FT /label= bar  
 FT /note= "coding region of bar gene of  
 FT Streptomyces hygroscopicus"  
 FT promoter complement (870..1702)  
 FT /\*tag= c  
 FT /label= P35S  
 FT /note= "35S promoter of Cauliflower Mosaic Virus"  
 FT complement (1740..2284)  
 FT /\*tag= d  
 FT /label= PTA29  
 FT /note= "promoter of TA29 gene of Nicotiana  
 FT tabacum"  
 FT CDS 2285..2557  
 FT /\*tag= e  
 FT /label= barstar  
 FT /note= "coding region of barstar gene of  
 FT Bacillus amyloliquefaciens"  
 FT misc\_signal 2558..2879  
 FT /\*tag= f  
 FT /label= 3'nos  
 FT /note= "3' regulatory sequence contg. the  
 FT polyadenylation signal of the nopaline  
 FT synthase gene of Agrobacterium T-DNA"  
 FT  
 FT misc\_RNA 1..38  
 FT /\*tag= g  
 FT /label= pUC19  
 FT /note= "pUC19 derived sequence"  
 FT misc\_RNA 2880..4986  
 FT /\*tag= h  
 FT /label= pUC19  
 FT /note= "pUC19 derived sequence"  
 XX  
 XX WO9534634-A2.  
 XX 21-DEC-1995.  
 XX  
 PF 06-JUN-1995; 95WO-EP02157.  
 XX  
 PR 06-JUN-1994; 94US-0254776.  
 XX  
 XX (PLBZ ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Krebbers E, Leemans J, Williams M;  
 XX  
 XX WPI; 1996-049664/05.  
 XX  
 XX Transgenic plants contg. male sterility and colour-linked restorer  
 PT genotypes - used for prodn. of male sterile seeds identifiable from  
 PT their colour, also new truncated anthocyanin regulatory genes and  
 PT aleurone specific promoters  
 XX  
 PS Disclosure; Page 68-71; 104pp; English.  
 XX  
 XX Shortened forms of anthocyanin (An) regulatory genes (ARG) are easier  
 CC

CC to manipulate than the complete gene and still provide an prodn.  
 CC The C1 gene (and the C1-S gene) can be considerably shortened  
 CC while still retaining, under appropriate conditions, its  
 CC capability of conditioning anthocyanin prodn. in the aleurone  
 CC of seeds of cereal plants such as corn. A pref. shortened C1 gene  
 CC is comprised in pCOL9 (see AAT08975). The full C1 gene sequence is  
 CC given in AAT08973.  
 CC A truncated B-peru gene (presumed sequence = AAT08674; actual  
 CC sequence = AAT08977) may also be used.  
 CC Plasmids pTS256 (AAT08976) and pTS200 (AAT17246) were used in the  
 CC construction of vectors comprising the C1 and B-peru genes  
 CC as well as male-sterility gene and a selectable marker gene.  
 SQ Sequence 4896 BP; 1252 A; 1146 C; 1186 G; 1312 T; 0 other;

Query Match 100.0%; Score 24; DB 17; Length 4896;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcagaagtatcagcgacctccacc 24  
 Db 2316 tcagaagtatcagcgacctccacc 2339

RESULT 13  
 AAX15631  
 ID AAX15631 standard; DNA; 4896 BP.  
 AC AAX15631;  
 XX  
 DT 07-MAY-1999 (first entry)  
 DE Nucleotide sequence of plasmid pTS256.  
 KW C1 gene; maize; male-sterile corn line; anthocyanin production;  
 KW pTS256; ds.  
 XX Synthetic.  
 XX  
 PN US5880331-A.  
 XX  
 PD 09-MAR-1999.  
 XX  
 PF 07-JUN-1995; 95US-0485139.  
 PR 07-JUN-1995; 95US-0485139.  
 PR 06-JUN-1994; 94US-0254776.  
 XX  
 PA (PLB2 ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Krebbers E, Leemans J, Williams M;  
 DR WPI; 1999-204053/17.  
 XX  
 PT Process for maintaining a male-sterile corn line - using male  
 PT sterile plants lacking functional regulatory gene for anthocyanin  
 PT production  
 XX  
 PS Example 2; Columns 31-36; 35pp; English.  
 XX

CC The present sequence represents the nucleotide sequence of plasmid  
 CC pTS256. It is used in the course of the invention. The specification  
 CC describes a process for maintaining a male-sterile corn line, using male  
 CC sterile parent plants lacking a functional gene for anthocyanin  
 CC production, and a maintainer corn line comprising male fertile parent  
 CC plants containing foreign DNA comprising a restorer gene and an active  
 CC regulatory protein gene. By using the anthocyanin gene, the colour of  
 CC the male-sterile plants will differ from that of the male fertile  
 CC plants. This will enable the seeds harvested from the plants to be  
 CC easily separated into those that will grow into male-fertile plants and  
 CC those that will grow into male-sterile plants.

SQ Sequence 4896 BP; 1251 A; 1147 C; 1186 G; 1312 T; 0 other;  
 Query Match 100.0%; Score 24; DB 20; Length 4896;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcagaagtatcagcgacctccacc 24  
 Db 2317 tcagaagtatcagcgacctccacc 2340

RESULT 14  
 AAT39339  
 ID AAT39339 standard; DNA; 5864 BP.  
 XX  
 AC AAT39339;  
 XX  
 DT 22-JAN-1997 (first entry)  
 DE Plasmid pTC0113 T-DNA used to obtain male sterile oilseed rape.  
 XX Plasmid pTC0113; male sterile; barnase; ribonuclease; barstar;  
 KW transgenic plant; oilseed rape; canole; Brassica napus; ds.  
 XX Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature complement (1..25)  
 FT /tag= a  
 FT /label= RB  
 FT /note= "right border of Agrobacterium T-DNA"  
 FT polyA\_signal complement (98..330)  
 FT /tag= b  
 FT /label= 3'g7  
 FT /note= "region containing polyA signal of gene 7  
 FT CDS complement (331..882)  
 FT /tag= c  
 FT /label= bar  
 FT /note= "region coding for phosphinothricin  
 FT promoter complement (883..2608)  
 FT /tag= d  
 FT /label= Psu  
 FT /note= "promoter of Arabidopsis Rubisco small  
 FT polyA\_signal complement (2659..3031)  
 FT /tag= e  
 FT /label= 3'nos  
 FT /note= "region containing polyA signal of nopaline  
 FT CDS complement (3032..3367)  
 FT /tag= f  
 FT /label= Barnase  
 FT /note= "Bacillus amyloliquefaciens barnase coding  
 FT promoter complement (3368..4877)  
 FT /tag= g  
 FT /label= pTA29  
 FT /note= "promoter of stamen-specific TA29 gene of  
 FT promoter 4924..5216  
 FT /tag= h  
 FT /label= Pnos  
 FT /note= "promoter of nopaline synthase gene of  
 FT CDS 5217..5489  
 FT /tag= i  
 FT /label= Barstar  
 FT /note= "region coding for barstar of Bacillus  
 FT polyA\_signal 5490..5765  
 FT amyloliquefaciens"



```
FT FT /*tag= j
FT FT /label= 3'g7
FT FT /note= "region containing polyA signal of gene 7
FT FT of Agrobacterium T-DNA"
FT FT complement (5840..5864)
FT FT /*tag= k
FT FT /label= LB
FT FT /note= "left border of Agrobacterium T-DNA"
FT FT
FT FT
FT FT
FT FT
FT FT W09626283-A1.
FT FT
FT FT
FT FT
FT FT 29-AUG-1996.
FT FT
FT FT 21-FEB-1996; 96WO-EP00722.
FT FT
FT FT 21-FEB-1995; 95EP-0400364.
FT FT
FT FT (PLBZ ) PLANT GENETIC SYSTEMS NV.
FT FT
FT FT Botterman J, Cornelissen M, Michiels F;
FT FT
FT FT WPI; 1996-402373/40.
FT FT
FT FT Prodn. of male sterile plants by transforming with a chimaeric
FT FT construct - comprising a male sterility DNA e.g. barnase and a
FT FT co-regulating gene, e.g. barstar, into the nuclear genome, useful
FT FT for generating hybrid cultivars
FT FT
FT FT Example 3; Page 33-3743-47; 56pp; English.
FT FT
FT FT Plasmid pTCO113 (AAT39339) is a T-DNA vector containing a bar gene
FT FT under control of the PSU promoter, a barnase gene under control
FT FT of the stamen-specific pTA29 promoter, and a barstar (co-regulatory)
FT FT gene under control of the Pnos promoter. 87% Of oilseed rape
FT FT plants regenerated after Agrobacterium-mediated transformation
FT FT using pTCO113 were male sterile. Barnase expression disturbed the
FT FT function of stamen cells leading to male sterility. Constitutive
FT FT expression of barstar counteracted any low level expression of
FT FT barnase in non-stamen tissue.
FT FT
FT FT SQ Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;
FT FT
FT FT
FT FT
FT FT
FT FT Query Match 100.0%; Score 24; DB 17; Length 5864;
FT FT Best Local Similarity 100.0%; Pred. No. 0.071;
FT FT Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT FT
FT FT Qy 1 tcagaagtatcagcgacctccacc 24
FT FT |||||||||
FT FT Db 5248 tcagaagtatcagcgacctccacc 5271
FT FT
FT FT
FT FT
FT FT RESULT 15
FT FT AAD06990
FT FT ID AAD06990 standard; DNA; 5865 BP.
FT FT
FT FT XX
FT FT AC AAD06990;
FT FT
FT FT XX
FT FT DT 06-AUG-2001 (first entry)
FT FT
FT FT DE Chimeric T-DNA of plasmid pTCO113.
FT FT
FT FT XX
FT FT KW T-DNA; plasmid pTCO113; transgenic Brassica plant; transformation event;
FT FT male-sterility gene; chimeric; tobacco; ds.
FT FT
FT FT XX
FT FT OS Chimeric - Streptomyces hygroscopicus.
FT FT OS Chimeric - Arabidopsis thaliana.
FT FT OS Chimeric - Bacillus amyloliquefaciens.
FT FT OS Chimeric - Nicotiana tabacum.
FT FT OS Chimeric - Agrobacterium tumefaciens.
FT FT OS Chimeric - Unidentified.
FT FT
FT FT XX
FT FT FH Key Location/Qualifiers
```

```
FT FT misc_feature
FT FT 1..25
FT FT /*tag= a
FT FT /note= "Right border repeat from the TL-DNA from
FT FT pTiB6S3"
FT FT 26..53
FT FT misc_feature
FT FT /*tag= b
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 54..90
FT FT misc_feature
FT FT /*tag= c
FT FT /note= "Residual sequence from the TL-DNA at the
FT FT right border repeat"
FT FT 98..309
FT FT 3'UTR
FT FT /*tag= d
FT FT /note= "The 3' untranslated end from the TL-DNA
FT FT gene 7 (3'g7) of pTiB6S3"
FT FT 310..331
FT FT misc_feature
FT FT /*tag= e
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 332..883
FT FT CDS
FT FT /*tag= f
FT FT /product= "Protein encoded by bialaphos resistance
FT FT gene (bar) of Streptomyces hygroscopicus"
FT FT 884..2609
FT FT promoter
FT FT /*tag= g
FT FT /note= "Promoter from the atSLA ribulose-1,5-biphosphate
FT FT carboxylase small subunit gene from Arabidopsis thaliana"
FT FT 2610..2659
FT FT misc_feature
FT FT /*tag= h
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 2660..2920
FT FT misc_feature
FT FT /*tag= i
FT FT /note= "TaqI fragment from the 3' untranslated end of the
FT FT nopaline synthase gene (3'nos) from the T-DNA of pTiT37
FT FT and containing plant polyadenylation signals"
FT FT 2921..2936
FT FT misc_feature
FT FT /*tag= j
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 2937..3032
FT FT 3'UTR
FT FT /*tag= k
FT FT /note= "The 3' untranslated region downstream from the
FT FT barnase coding sequence of Bacillus amyloliquefaciens"
FT FT 3033..3368
FT FT CDS
FT FT /*tag= l
FT FT /product= "Protein encoded by barnase gene from
FT FT Bacillus amyloliquefaciens"
FT FT 3369..4878
FT FT promoter
FT FT /*tag= m
FT FT /note= "Promoter region of the anther-specific gene
FT FT TA29 from Nicotiana tabacum"
FT FT 4879..4924
FT FT misc_feature
FT FT /*tag= n
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 4925..5215
FT FT promoter
FT FT /*tag= o
FT FT /note= "Promoter of the nopaline synthase gene from the
FT FT T-DNA of pTiT37 of Agrobacterium tumefaciens"
FT FT 5216..5217
FT FT misc_feature
FT FT /*tag= p
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 5218..5490
FT FT CDS
FT FT /*tag= q
FT FT /product= "Protein encoded by barstar gene of
FT FT Bacillus amyloliquefaciens"
FT FT 5491..5530
FT FT misc_feature
FT FT /*tag= r
FT FT /note= "Sequence from the 3' untranslated end of the
FT FT barstar gene from Bacillus amyloliquefaciens"
FT FT 5531..5554
FT FT misc_feature
FT FT /*tag= s
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 5555..5766
FT FT 3'UTR
FT FT /*tag= t
FT FT /note= "The 3' untranslated end from the TL-DNA
FT FT
```

```
FT misc_feature gene 7 (3'g7) of pTiB6S3"
FT 5767..5773 /*tag= u
FT /*note= "Synthetic polylinker derived sequence"
FT 5774..5810
FT /*tag= v
FT /*note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 5811..5840
FT /*tag= w
FT /*note= "Synthetic polylinker derived sequence"
FT 5841..5865
FT /*tag= x
FT /*note= "Left border repeat from the TL-DNA from
FT pTiB6S3"
XX
XX WO200131042-A2.
XX
XX 03-MAY-2001.
XX
XX 26-OCT-2000; 2000WO-EP10680.
XX
XX 29-OCT-1999; 99US-0430497.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX Weston B, De Beuckeleer M;
XX WPI; 2001-300517/31.
XX
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX harboring specific transformation events, particularly by presence of
XX male-sterility gene, at specific location in its genome -
XX
XX Claim 1; Page 47-49; 53pp; English.
XX
XX The present invention relates to a transgenic Brassica plant or its
XX seed, cells or tissues, characterised by harbouring a specific
XX transformation event, particularly by the presence of a male-sterility
XX gene, at a specific location in the Brassica genome. Transgenic
XX Brassica plant is useful for producing a hybrid seed by crossing the
XX transgenic plant with a male-fertile Brassica plant and harvesting the
XX hybrid seed from the transgenic Brassica plant.
XX The present sequence is chimeric T-DNA of plasmid pTCO113. This sequence
XX comprises right border repeat, left border repeat and 3' untranslated
XX region (UTR) from TL-DNA of pTiB6S3, synthetic polylinker sequences,
XX coding regions of blattaphos resistance gene (bar) from
XX Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens
XX and barstar gene from Bacillus subtilis and promoters of atS1A
XX ribulose-1,5-bisphosphate carboxylase small subunit gene from
XX Arabidopsis thaliana, the anther-specific gene TA29 from
XX Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTi37
XX of Agrobacterium tumefaciens.
XX
XX Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;
```

Query Match 100.0%; Score 24; DB 22; Length 5865;  
Best Local Similarity 100.0%; Pred. NO. 0.071;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 5249 tcagaagtatcagcgacctccacc 5272

Search completed: February 25, 2002, 18:17:36  
Job time: 16694 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:27 ; Search time 301.6 Seconds  
(without alignments)  
18.022 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24

Sequence: 1 tcagaagtatcagcgacctccacc 24

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	24	100.0	1303	US-08-894-440-2	Sequence 2, Appli
2	24	100.0	3544	US-08-485-139-3	Sequence 3, Appli
3	24	100.0	3544	US-08-750-357-3	Sequence 3, Appli
4	24	100.0	4808	US-08-351-413-17	Sequence 17, Appli
5	24	100.0	4808	US-09-025-583-17	Sequence 17, Appli
6	24	100.0	4896	US-08-485-139-2	Sequence 2, Appli
7	24	100.0	4896	US-08-750-357-2	Sequence 2, Appli
8	24	100.0	5864	US-08-894-440-4	Sequence 4, Appli
9	24	100.0	6555	US-08-351-413-2	Sequence 2, Appli
10	24	100.0	6555	US-09-025-583-2	Sequence 2, Appli
11	16.6	69.2	1866	US-08-403-388-1	Sequence 1, Appli
12	16.6	69.2	1866	US-08-658-578-1	Sequence 1, Appli
13	16.6	69.2	1866	US-08-846-111D-1	Sequence 1, Appli
14	16.6	69.2	1866	US-09-056-105-19	Sequence 19, Appli
15	16.6	69.2	40352	US-08-846-111D-15	Sequence 15, Appli
16	16	66.7	1128	US-09-016-366A-20	Sequence 20, Appli
17	16	66.7	1128	US-08-978-404B-15	Sequence 15, Appli
18	16	66.7	1137	US-09-016-366A-18	Sequence 18, Appli
19	16	66.7	1137	US-08-978-404B-13	Sequence 13, Appli
20	16	66.7	1154	US-09-016-366A-16	Sequence 16, Appli
21	16	66.7	1154	US-08-978-404B-11	Sequence 11, Appli
22	15.6	65.0	4286	US-09-413-304-7	Sequence 7, Appli
23	15.6	65.0	4437	US-08-559-303B-72	Sequence 72, Appli
24	15.6	65.0	4437	US-09-175-828-72	Sequence 72, Appli
25	15.6	65.0	6519	US-08-588-985-1	Sequence 1, Appli
26	15.6	65.0	6519	US-08-971-988-1	Sequence 1, Appli
27	15.2	63.3	363	US-08-594-031-104	Sequence 104, App

c	28	15.2	63.3	1082	1	US-08-716-301-5	Sequence 5, Appli
	29	15.2	63.3	2326	2	US-08-231-193A-41	Sequence 41, Appli
	30	15.2	63.3	2326	2	US-08-486-273A-41	Sequence 41, Appli
	31	15.2	63.3	2326	3	US-08-480-474-41	Sequence 41, Appli
	32	15.2	63.3	2326	3	US-08-940-086A-41	Sequence 41, Appli
	33	15.2	63.3	3243	2	US-08-231-193A-44	Sequence 44, Appli
	34	15.2	63.3	3243	2	US-08-486-273A-44	Sequence 44, Appli
	35	15.2	63.3	3243	3	US-08-480-474-44	Sequence 44, Appli
	36	15.2	63.3	3243	3	US-08-940-086A-44	Sequence 44, Appli
	37	15.2	63.3	3698	2	US-08-231-193A-43	Sequence 43, Appli
	38	15.2	63.3	3698	2	US-08-486-273A-43	Sequence 43, Appli
	39	15.2	63.3	3698	3	US-08-480-474-43	Sequence 43, Appli
	40	15.2	63.3	3698	3	US-08-940-086A-43	Sequence 43, Appli
	41	15.2	63.3	3717	1	US-08-026-138E-7	Sequence 7, Appli
	42	15.2	63.3	4002	2	US-08-231-193A-53	Sequence 53, Appli
	43	15.2	63.3	4002	2	US-08-486-273A-53	Sequence 53, Appli
	44	15.2	63.3	4002	3	US-08-480-474-53	Sequence 53, Appli
	45	15.2	63.3	4002	3	US-08-940-086A-53	Sequence 53, Appli

ALIGNMENTS

RESULT 1  
US-08-894-440-2  
; Sequence 2, Application US/08894440  
; Patent NO. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894.440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1303  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI  
; OTHER INFORMATION: fragment of pTS88  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(35)  
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36)..(694)  
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain  
; OTHER INFORMATION: CM1841 (p35S)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (695)..(967)  
; OTHER INFORMATION: region coding for barstar of Bacillus  
; OTHER INFORMATION: amyloliquefians  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (968)..(1287)  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1288)..(1303)  
; OTHER INFORMATION: polylinker of pGEM2  
US-08-894-440-2

Query Match 100.0%; Score 24; DB 3; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 0.0042;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 tcagaagtatcagcgacctccacc 24

```
Db 726 tcagaagtatcagcgacctccacc 749
|||||
RESULT 2
US-08-485-139-3/C
; Sequence 3, Application US/08485139
; Patent No. 5880331
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,139
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3544 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: EcoRI-HindIII region of plasmid pTS200
; FEATURE:
; NAME/KEY: -
; LOCATION: 3227..3504
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2675..3226
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces
; OTHER INFORMATION: hygroscopicus"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1841..2674
; OTHER INFORMATION: /label= p35S
; OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (626..1803)
; OTHER INFORMATION: /label= PCA55
; OTHER INFORMATION: /note= "promoter of CA55 gene of Zea mays"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (353..625)
```

```
;
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of barstar gene of Bacillus
; OTHER INFORMATION: amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (30..352)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..6
; OTHER INFORMATION: /label= EcoRI
; FEATURE:
; NAME/KEY: -
; LOCATION: 3539..3544
; OTHER INFORMATION: /label= HindIII
; US-08-485-139-3
Query Match 100.0%; Score 24; DB 2; Length 3544;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcagaagtatcagcgacctccacc 24
|||||
Db 594 TCAGAAGTATCAGCGACCTCCACC 571
|||||
RESULT 3
US-08-750-357-3/C
; Sequence 3, Application US/08750357
; Patent No. 6008437
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,357
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 018030-100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3544 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: EcoRI-HindIII region of plasmid pTS200
```

```
FEATURE:
NAME/KEY: 3227..3504
LOCATION: 3227..3504
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"
FEATURE:
NAME/KEY: 2675..3226
LOCATION: 2675..3226
OTHER INFORMATION: /label= bar
OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces hygroscopicus"
FEATURE:
NAME/KEY: 1841..2674
LOCATION: 1841..2674
OTHER INFORMATION: /label= P35S
OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
FEATURE:
NAME/KEY: 626..1803
LOCATION: 626..1803
OTHER INFORMATION: /label= PCA55
OTHER INFORMATION: /note= "promoter of CA55 gene of Zea mays"
FEATURE:
NAME/KEY: 353..625
LOCATION: 353..625
OTHER INFORMATION: /label= barstar
OTHER INFORMATION: /note= "coding region of barstar gene of Bacillus"
OTHER INFORMATION: amyloliquefaciens"
FEATURE:
NAME/KEY: 30..352
LOCATION: 30..352
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"
FEATURE:
NAME/KEY: 1..6
LOCATION: 1..6
OTHER INFORMATION: /label= EcoRI
OTHER INFORMATION:
FEATURE:
NAME/KEY: 3539..3544
LOCATION: 3539..3544
OTHER INFORMATION: /label= HindIII
US-08-750-357-3

Query Match 100.0%; Score 24; DB 3; Length 3544;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcagctccacc 24
|||||
Db 594 TCAGAAGTATCAGCAGCTCCACC 571

RESULT 4
US-08-351-413-17/c
; Sequence 17, Application US/08351413
; Patent No. 5750867
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/351,413
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/899,072
FILING DATE: 12-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,849
FILING DATE: 03-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 2121-102PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 4808 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: EcoRI-HindIII fragment of plasmid pYS218
FEATURE:
NAME/KEY: -
LOCATION: complement (18..401)
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the
OTHER INFORMATION: polyadenylation site derived from Agrobacterium
OTHER INFORMATION: T-DNA nopaline synthase gene"
FEATURE:
NAME/KEY: -
LOCATION: complement (402..737)
OTHER INFORMATION: /label= barnase
OTHER INFORMATION: /note= "coding region of the barnase gene of
OTHER INFORMATION: Bacillus amyloliquefaciens"
FEATURE:
NAME/KEY: -
LOCATION: complement (738..1944)
OTHER INFORMATION: /label= PZM13
OTHER INFORMATION: /note= "promoter region of the Zm13 gene of Zea
OTHER INFORMATION: mays"
FEATURE:
NAME/KEY: -
LOCATION: complement (1945..2281)
OTHER INFORMATION: /label= 3'nos
FEATURE:
NAME/KEY: -
LOCATION: complement (2282..2554)
OTHER INFORMATION: /label= barstar
OTHER INFORMATION: /note= "coding region of the barstar gene of
OTHER INFORMATION: Bacillus amyloliquefaciens"
FEATURE:
NAME/KEY: -
LOCATION: complement (2555..3099)
OTHER INFORMATION: /label= PTA29
OTHER INFORMATION: /note= "promoter region of the TA29 gene of
OTHER INFORMATION: Nicotiana tabacum"
FEATURE:
NAME/KEY: -
LOCATION: 3100..3932
OTHER INFORMATION: /label= 35S3
OTHER INFORMATION: /note= "35S3" promoter sequence derived from
OTHER INFORMATION: cauliflower mosaic virus isolate CabbB-J1"
FEATURE:
NAME/KEY: -
```

```

; LOCATION: 3933..4484
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of the phosphinothricin
; OTHER INFORMATION: acetyltransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4485..4763
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: 2333..2356
; OTHER INFORMATION: /label= BXOL2
; OTHER INFORMATION: /note= "region corresponding to oligonucleotide
; OTHER INFORMATION: BXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2538..2586)
; OTHER INFORMATION: /label= PTA29SBOXL2
; OTHER INFORMATION: /note= "region complementary to oligonucleotide
; OTHER INFORMATION: PTA29SBOXL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2800..2823)
; OTHER INFORMATION: /label= PTA29OL5
; OTHER INFORMATION: /note= "region complementary to part of
; OTHER INFORMATION: oligonucleotide PTA29OL5"
; US-08-351-413-17
;
; Query Match 100.0%; Score 24; DB 1; Length 4808;
; Best Local Similarity 100.0%; Pred. No. 0.0054;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 tcgaagatcagcagcctccacc 24
; Db 2529 TCAGAAGTATCAGCGCCTCCACC 2506
;
; RESULT 5
; US-09-025-583-17/c
; Sequence 17, Application US/09025583
; Patent No. 5977433
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,583
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.

```

```

; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: EcoRI-HindIII fragment of plasmid pTS218
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (18..401)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (402..737)
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding region of the barnase gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (738..1944)
; OTHER INFORMATION: /label= PZM13
; OTHER INFORMATION: /note= "promoter region of the Zml3 gene of Zea
; OTHER INFORMATION: mays"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (1945..2281)
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2282..2554)
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of the barstar gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2555..3099)
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "promoter region of the TA29 gene of
; OTHER INFORMATION: Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3100..3932
; OTHER INFORMATION: /label= 35S3
; OTHER INFORMATION: /note= "35S3" promoter sequence derived from
; OTHER INFORMATION: cauliflower mosaic virus isolate cabBB-J1"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3933..4484
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of the phosphinothricin
; OTHER INFORMATION: acetyltransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4485..4763
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: 2333..2356
; OTHER INFORMATION: /label= BXOL2
; OTHER INFORMATION: /note= "region corresponding to oligonucleotide

```

```

; OTHER INFORMATION: BXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2538..2586)
; OTHER INFORMATION: /label= TA29SBXOL2
; OTHER INFORMATION: /note= "region complementary to oligonucleotide
; OTHER INFORMATION: TA29SBXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2800..2823)
; OTHER INFORMATION: /label= PTA29OL5
; OTHER INFORMATION: /note= "region complementary to part of
; OTHER INFORMATION: oligonucleotide PTA29OL5"
US-09-025-583-17
```

```

Query Match 100.0%; Score 24; DB 2; Length 4808;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 tcagaagtatcagcgacctccacc 24
|||||
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506
```

```

RESULT 6
US-08-485-139-2
; Sequence 2, Application US/08485139
; Patent No. 5880331
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,139
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid pTS256, linearized at HindIII
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (39..317)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
```

```

; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (318..869)
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces
; OTHER INFORMATION: hygroscopicus"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (870..1702)
; OTHER INFORMATION: /label= P35S
; OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1740..2284
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "promoter of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2285..2557
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of barstar gene of Bacillus
; OTHER INFORMATION: amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2558..2879
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..38
; OTHER INFORMATION: /label= pUC19
; OTHER INFORMATION: /note= "pUC19 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2880..4896
; OTHER INFORMATION: /label= pUC19
; OTHER INFORMATION: /note= "pUC19 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3004..3009
; OTHER INFORMATION: /label= EcoRI
US-08-485-139-2
```

```

Query Match 100.0%; Score 24; DB 2; Length 4896;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 tcagaagtatcagcgacctccacc 24
|||||
Db 2317 TCAGAAGTATCAGCGACCTCCACC 2340
```

```

RESULT 7
US-08-750-357-2
; Sequence 2, Application US/08750357
; Patent No. 6008437
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
```

```
STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,357
FILING DATE: 21-MAR-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 018030-100
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4896 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ORGANISM: plasmid pTS256, linearized at HindIII
FEATURE:
NAME/KEY: -
LOCATION: 39..317
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"
FEATURE:
NAME/KEY: -
LOCATION: 318..869
OTHER INFORMATION: /label= bar
OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces hygroscopicus"
FEATURE:
NAME/KEY: -
LOCATION: 870..1702
OTHER INFORMATION: /label= P35S
OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
FEATURE:
NAME/KEY: -
LOCATION: 1740..2284
OTHER INFORMATION: /label= PTA29
OTHER INFORMATION: /note= "promoter of TA29 gene of Nicotiana tabacum"
FEATURE:
NAME/KEY: -
LOCATION: 2285..2557
OTHER INFORMATION: /label= barstar
OTHER INFORMATION: /note= "coding region of barstar gene of
OTHER INFORMATION: Bacillus amyloliquefaciens"
FEATURE:
NAME/KEY: -
LOCATION: 2558..2879
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"
FEATURE:
NAME/KEY: -
LOCATION: 1..38
OTHER INFORMATION: /label= pUC19
OTHER INFORMATION: /note= "pUC19 derived sequence"
FEATURE:
NAME/KEY: -
LOCATION: 2880..4896
OTHER INFORMATION: /label= pUC19
OTHER INFORMATION: /note= "pUC19 derived sequence"
FEATURE:
NAME/KEY: -
```

```
LOCATION: 3004..3009
OTHER INFORMATION: /label= ECORI
US-08-750-357-2

Query Match
Best Local Similarity 100.0%; Score 24; DB 3; Length 4896;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24
|||||
Db 2316 TCAGAAATATCAGCGACCTCCACC 2339
```

```
RESULT 8
US-08-894-440-4
Sequence 4, Application US/08894440
Patent No. 6025546
GENERAL INFORMATION:
APPLICANT: PLANT GENETIC SYSTEMS N.V.
TITLE OF INVENTION: Method to obtain male sterile plants
FILE REFERENCE: NMSCOR
CURRENT APPLICATION NUMBER: US/08/894,440
CURRENT FILING DATE: 1997-11-12
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 5864
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
OTHER INFORMATION: plasmid pTC0113
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement(1)..(25))
OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((98)..(330))
OTHER INFORMATION: region containing polyadenylation signal of gene 7
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((331)..(882))
OTHER INFORMATION: region coding for phosphinothricin acetyl
OTHER INFORMATION: transferase (bar)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((883)..(2608))
OTHER INFORMATION: promoter of small subunit gene of Rubisco of
OTHER INFORMATION: Arabidopsis (Pssu)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((2659)..(3031))
OTHER INFORMATION: region containing polyadenylation signal of
OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((3032)..(3367))
OTHER INFORMATION: region coding for barnase of Bacillus
OTHER INFORMATION: amyloliquefaciens
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((3368)..(4877))
OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
OTHER INFORMATION: tabacum (PTA29)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4924)..(5216)
OTHER INFORMATION: promoter of nopaline synthase gene of
OTHER INFORMATION: Agrobacterium T-DNA (Phos)
```



FEATURE: NO  
NAME/KEY: misc\_feature  
LOCATION: (5217)..(5489)  
OTHER INFORMATION: region coding for barstar of Bacillus  
OTHER INFORMATION: amyloliquefaciens

FEATURE: NO  
NAME/KEY: misc\_feature  
LOCATION: (5490)..(5765)  
OTHER INFORMATION: region containing polyadenylation signal of gene 7  
OTHER INFORMATION: T-DNA (3'g7)  
FEATURE: NO  
NAME/KEY: misc\_feature  
LOCATION: Complement((5840)..(5864))  
OTHER INFORMATION: left border of Agrobacterium T-DNA  
US-08-894-440-4

Query Match 100.0%; Score 24; DB 3; Length 5864;  
Best Local Similarity 100.0%; Pred. No. 0.0056;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcagcctccacc 24  
|||||

Db 5248 tcagaagtatcagcagcctccacc 5271

## RESULT 9

US-08-351-413-2/c  
Sequence 2, Application US/08351413  
Patent No. 5750867

GENERAL INFORMATION:  
APPLICANT: Williams, Mark  
APPLICANT: Leemans, Jan  
TITLE OF INVENTION: Maintenance of male-sterile plants  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 2046

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA: US/08/351.413

FILING DATE:

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/899,072

FILING DATE: 12-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/970,849

FILING DATE: 03-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: Svensson, Leonard R.

REGISTRATION NUMBER: 30,330

REFERENCE/DOCKET NUMBER: 2121-102PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 6555 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO  
ORIGINAL SOURCE: Plasmid pVE144 (replicable in E.coli)  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: 1..396  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: complement (397..751)  
OTHER INFORMATION: /label= 3'nos  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA nopaline synthase gene"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: complement (752..1024)  
OTHER INFORMATION: /label= barstar  
OTHER INFORMATION: /note= "coding region of the barstar gene of  
OTHER INFORMATION: Bacillus amyloliquefaciens"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: complement (1025..1607)  
OTHER INFORMATION: /label= TA29  
OTHER INFORMATION: /note= "promoter derived from the TA29 gene of  
OTHER INFORMATION: Nicotiana tabacum"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: 1608..2440  
OTHER INFORMATION: /label= 35S3  
OTHER INFORMATION: /note= "35S3 promoter sequence derived from  
OTHER INFORMATION: cauliflower mosaic virus isolate Cabb8-JI"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: 2441..3256  
OTHER INFORMATION: /label= neo  
OTHER INFORMATION: /note= "coding region of the neomycin  
OTHER INFORMATION: phosphotransferase gene of Tn5"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: 3257..4315  
OTHER INFORMATION: /label= 3'ocs  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA octopine synthase gene"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: 4316..6555  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"

US-08-351-413-2

Query Match 100.0%; Score 24; DB 1; Length 6555;  
Best Local Similarity 100.0%; Pred. No. 0.0057;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcagcctccacc 24

|||||

Db 993 TCAGAAGTATCAGCGACCTCCACC 970

## RESULT 10

US-09-025-583-2/c  
Sequence 2, Application US/09025583  
Patent No. 5977433

GENERAL INFORMATION:

APPLICANT: Williams, Mark

APPLICANT: Leemans, Jan

TITLE OF INVENTION: Maintenance of male-sterile plants

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/09/025,583
; APPLICATION NUMBER: US/09/025,583
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6555 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: plasmid pVEL44 (replicable in E.coli)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..396
; OTHER INFORMATION: /label= pUC18
; OTHER INFORMATION: /note= "pUC18 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (397..751)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (752..1024)
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of the barstar gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (1025..1607)
; OTHER INFORMATION: /label= TA29
; OTHER INFORMATION: /note= "promoter derived from the TA29 gene of
; OTHER INFORMATION: Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1608..2440
; OTHER INFORMATION: /label= 35S3
; OTHER INFORMATION: /note= "35S3 promoter sequence derived from
; OTHER INFORMATION: cauliflower mosaic virus isolate Cabbb-J1"
; FEATURE:

```

```

; NAME/KEY: -
; LOCATION: 2441..3256
; OTHER INFORMATION: /label= neo
; OTHER INFORMATION: /note= "coding region of the neomycine
; OTHER INFORMATION: phosphotransferase gene of Tn5"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3257..4315
; OTHER INFORMATION: /label= 3'ocs
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA octopine synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4316..6555
; OTHER INFORMATION: /label= pUC18
; OTHER INFORMATION: /note= "pUC18 derived sequence"
; US-09-025-583-2
;
; Query Match 100.0%; Score 24; DB 2; Length 6555;
; Best Local Similarity 100.0%; Pred. No. 0.0057;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 tcagaagtatcagcgacctccacc 24
; | | | | | | | | | | | | | | | | | |
; Db 993 TCAGAGTATCAGCGACCTCCACC 970
;
; RESULT 11
; US-08-403-388-1
; Sequence 1, Application US/08403388
; Patent No. 5587289
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members Of The Mage-Xp Family and Uses Thereof
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,388
; FILING DATE: 14-MARCH-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587289man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1866 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-403-388-1
;
; Query Match 69.2%; Score 16.6; DB 1; Length 1866;
; Best Local Similarity 82.6%; Pred. No. 21;
; Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 cagaagtatcagcgacctccacc 24
   ||||| |||| || |||||
DB 498 CAGAAAGCCTCAGGAGCTCCACC 520

RESULT 12
US-08-658-578-1
; Sequence 1, Application US/08658578
; Patent No. 5759783
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members Of The Mage-Xp Family and Uses Thereof
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,578
; FILING DATE: 5-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,388
; FILING DATE: 14-MARCH-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5759783man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5444
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1866 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-658-578-1

Query Match 69.2%; Score 16.6; DB 1; Length 1866;
Best Local Similarity 82.6%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cagaagtatcagcgacctccacc 24
   ||||| |||| || |||||
DB 498 CAGAAAGCCTCAGGAGCTCCACC 520

RESULT 13
US-08-846-111D-1
; Sequence 1, Application US/08846111D
; Patent No. 6017705
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members Of The Mage-B Family and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York

```

```

; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect For DOS 6.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,111D
; FILING DATE: 25-APRIL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/658,578
; FILING DATE: 5-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,388
; FILING DATE: 14-MARCH-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6017705man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5444.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1866 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-846-111D-1

Query Match 69.2%; Score 16.6; DB 3; Length 1866;
Best Local Similarity 82.6%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cagaagtatcagcgacctccacc 24
   ||||| |||| || |||||
DB 498 CAGAAAGCCTCAGGAGCTCCACC 520

RESULT 14
US-09-056-105-19
; Sequence 19, Application US/09056105
; Patent No. 6287569
; GENERAL INFORMATION:
; APPLICANT: KIPPS, THOMAS J.
; APPLICANT: WU, YUNQI
; TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR
; TITLE OF INVENTION: PROCESSING
; FILE REFERENCE: 233/221
; CURRENT APPLICATION NUMBER: US/09/056,105
; CURRENT FILING DATE: 1998-04-06
; EARLIER APPLICATION NUMBER: 60/043,467
; EARLIER FILING DATE: 1997-04-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-056-105-19

Query Match 69.2%; Score 16.6; DB 4; Length 1866;
Best Local Similarity 82.6%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cagaagtatcagcgacctccacc 24
   ||||| |||| || |||||
DB 498 cagaagctcaggagacctccacc 520

```

```

RESULT 15
US-08-846-111D-15
; Sequence 15, Application US/08846111D
; Patent No. 6017705
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members Of The Mage-B Family and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect For DOS 6.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,111D
; FILING DATE: 25-APRIL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/658,578
; FILING DATE: 5-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,388
; FILING DATE: 14-MARCH-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6017705man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5444.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40352 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-846-111D-15

```

```

Query Match      69.2%; Score 16.6; DB 3; Length 40352;
Best Local Similarity 82.6%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24
||||| ||||| ||||| |||||
Db 38322 CAGAGGCTCAGGAGCTCCACC 38344

```

Search completed: February 25, 2002, 18:05:29  
Job time: 18577 sec